الإحصاء
فى البحوث العلمية

محمد أبويوسف

استاذ الإحصاء بكلية التربية
جامعة عين شمس

الناشر



المكتبة الأكاديمية

الإحصـــــاء
فى البحوث العلمية

اهداءات ٢٠٠١
صيدلي/ حسن سعد الدين حجازي
الإسكندرية

الإحصـــــاء
فى البحوث العلمية

محمد أبويوسف

أستاذ الإحصاء بكلية التربية
جامعة عين شمس

الناشر



المكتبة الأكاديمية

١٢١ ش التحرير ـ الدقى ـ القاهرة ت : ٣٤٩١٨٩٦ / ٣٤٨٥٢٨٢ تلكس : ٩٤١٢٤ ABCMN UN

بسم الله الرحمن الرحيم

قُلْ إِنَّ صَلَاتِي وَنُسُكِي وَمَحْيَايَ وَمَمَاتِي لِلَّهِ رَبِّ الْعَالَمِينَ

صدق الله العظيم

بسم الله الرحمن الرحيم

مقدمة

يتناول هذا الكتاب أسس التحليل الإحصائي للتجارب الميدانية والمعملية في مجال العلوم . ولما كانت الدراسة المثمرة للإحصاء التطبيقي تتطلب حدا أدنى من المعرفة بالنظرية الإحصائية ، وإغفال ذلك يؤدي إلى فهم سطحي تنجم عنه أخطاء جسيمة في التطبيق العملي ، فقد عنى الكتاب بإرساء أساسيات وركائز هذه النظرية كما عنى بالربط بين النظرية والتطبيق وتقديم الأصول العلمية لشروط ومحددات مايستخدم من طرق واختبارات وعمليات استدلال . وقد تجنب في ذلك كله الدخول في التفاصيل والبراهين الرياضية التي قد تعوق القارىء عن متابعة مسيرة التسلسل المنطقي للمفاهيم والأساليب الرئيسية المنشودة .

وقراءة هذا الكتاب لا تتطلب إلا اليسير من الخلفية الرياضية التي لا تخرج عن المبادىء الحسابية والجبرية البسيطة . وفي الموضوعات التي تحتاج إلى جهد كبير في إجراء العمليات الحسابية اللازمة للتحليل اعتمد الكتاب على الحاسب الالكتروني توفيرا للجهد وضمانا للدقة والسرعة ، دون أن يستلزم ذلك أن يكون الباحث قادرا على تشغيل الحاسب بنفسه بل يكفيه الاتصال بأحد مراكز الحاسب التي أصبحت اليوم متوافرة في الجامعات ومراكز البحث العلمي .

ولابد من الإشارة هنا إلى أن هذا الكتاب هو في حقيقة أمره طبعة مزيدة من كتاب سبق لي تأليفه بعنوان « مقدمة في الإحصاء البيولوجي » تكفلت جامعة

قطر مشكورة بإصداره سنة ١٩٨٤ مساهمة منها في المد الثقافى العربى وتبادل الكتب والمطبوعات العلمية مع الجامعات العربية والهيئات الثقافية الأخرى . ولقد سعدت بتدريس ذلك الكتاب عدة سنوات خلال فترة عملى بكلية العلوم بتلك الجامعة لطلاب من مختلف الشعب العلمية فى مرحلة البكالوريوس وفى الدراسات العليا التمهيدية . ولقد ضاعفت الزيادة في هذا الكتاب من حجم الكتاب السابق ، وهى تتمثل فى توسيع بعض الفصول خاصة فصلى تحليل التباين والانحدار الخطى البسيط ، وكذلك فى إضافة خمسة فصول جديدة تحمل طابعا متقدما هى الفصول السابع والحادى عشر والثانى عشر والثالث عشر والخامس عشر . وتستهدف هذه الزيادة استكمال الجوانب التطبيقية لعلم الإحصاء تعزيزا لمحتوى الكتاب بما يخدم احتياجات قطاع أكبر من الطلاب والباحثين ، ومساهمة فى سد إحدى الثغرات العديدة التى تعانى منها المكتبة العربية في مجال العلوم .

ويرجو المؤلف أن يكون في الأسلوب الذى قدم به المادة وما استخدمه من أمثلة مستمد أغلبها من بحوث ودراسات واقعية مايمكن القارىء من اكتساب منهج فكرى في التحليل الإحصائى يجعله قادرا على المساهمة فى تخطيط التجارب وتحليلها ، والاعتماد على نفسه فى الاستزادة من دراسة الإحصاء التطبيقى ، وفى تفهم ما ينشر من البحوث فى هذا الميدان . ولعل فى تقديم المصطلحات الإحصائية باللغتين العربية والانجليزية ما ييسر للقارىء متابعة المراجع والبحوث الأجنبية التى لا غنى للباحث عنها .

أسأل الله التوفيق وعلى الله قصد السبيل .

المؤلف

محمد أبو يوسف

١٩٨٩/٨/١٨

المحتويات

الصفحات

الفصل الأول : مفاهيم أولية .... ١٥ – ٣٩

المجتمع والعينة – العينة العشوائية البسيطة – المتغيرات الإحصائية – الأخطاء العشوائية – قواعد تقريب الأعداد – الاحتمال – النماذج الرياضية .

الفصل الثانى : التوزيعات التكرارية .... ٤١ – ٨٥

الجداول التكرارية – التمثيل البيانى للتوزيعات التكرارية – المئينات والربيعات – الوصف العددى للتوزيعات التكرارية (الوسط الحسابى والانحراف المعيارى ، الالتواء ، التفرطح) – شكل الساق والورقة .

الفصل الثالث : بعض نماذج الاحتمال .... ٨٧ – ١٢١

توزيعات الاحتمال الوثابة – نماذج الاحتمال – توزيع ذى الحدين – تقدير الدليل ح – توفيق توزيع ذى الحدين لتوزيع تكرارى معلوم – توزيع بواسون (كتقريب لتوزيع ذى الحدين – كنموذج لتوزيع الأحداث النادرة) – اختبار استقلال الأحداث النادرة – نمط التجمع ونمط التنافر – توزيع باسكال – التوزيع الهندسى – توزيعات الاحتمال المتصلة .

الفصل الرابع : التوزيع المعتدل والتوزيع المعتدل اللوغاريتمي .... ١٢٣ – ١٤٧

(أولا) : التوزيع المعتدل – بعض خواص التوزيع – جداول

المساحات أسفل المنحنى المعتدل المعيارى – الكشف عن الاعتدالية – طريقة بيانية للكشف عن الاعتدالية – معالجة عدم اعتدالية التوزيع – تقريب توزيع ذى الحدين بتوزيع معتدل .

(ثانيا) : التوزيع المعتدل اللوغاريتمى – بعض خواصه .

الفصل الخامس : توزيعات خاصة .... ١٤٩ – ١٥٨

توزيعات ت ، χ^2 ، ف وجداول القيم الحرجة .

الفصل السادس : نظرية العينات .... ١٥٩ – ٢٢٦

نظرية العينات – توزيعات المعاينة – الإحصاءة – الخطأ المعيارى – التقدير غير المتحيز – المعاينة من مجتمعات معتدلة – المعاينة من توزيع ذى الحدين – اختبارات الفروض – اختبار ت (اختبار فرض عن الوسط الحسابى لمجتمع معتدل – فترات الثقة للوسط الحسابى لمجتمع معتدل – مقارنة متوسطى مجتمعين معتدلين – اختبار استقلال الأحداث النادرة) – اختبار χ^2 (اختبار فرض عن توزيع مجتمع – اختبار حسن المطابقة – اختبار استقلال خاصتين – فترات الثقة لتباين مجتمع معتدل) – اختبار ف – فترات الثقة للنسبة في مجتمع وللفرق بين نسبتين – تحديد حجم العينة – مراقبة الإنتاج .

الفصل السابع : حساسية اختبارات الفروض .... ٢٢٧ – ٢٥٨

نوعا الأخطاء الإحصائية – حساب قيمة β فى اختبار ذى جانب واحد عن متوسط مجتمع معتدل (دالة القوة – العوامل المؤثرة فى الخطأ من النوع الثانى – كيفية زيادة قوة الاختبار) – حساب قيمة β فى اختبار ذى جانبين عن متوسط مجتمع معتدل – حساب قيمة β فى اختبار الفرق بين مجتمعين معتدلين – حساب قيمة β فى اختبار النسبة .

الفصل الثامن : تحليل التباين وتصميم التجارب .... ٢٥٩ – ٣٦١
التحليل الإحصائى وتصميم التجارب – تحليل التباين – مصطلحات وتعاريف – التجارب ذوات العامل الواحد – المقارنة القبلية والبعدية بين المتوسطات – التجارب ذوات العاملين المتفاعلين وغير المتفاعلين – المقارنات التزاوجية – التجارب ذوات الثلاثة العوامل (فى مربع لاتينى) – معالجة الانحرافات عن افتراضات التحليل – عودة إلى مقارنة المتوسطات – المقارنات القبلية – معيار استقلال مقارنتين – اختبار المقارنات القبلية – تجزىء مجموع المربعات – اختبار المقارنات البعدية – النموذج عشوائى التأثيرات .

الفصل التاسع : الانحدار الخطى البسيط .... ٣٦٣ – ٤١٢
المجتمع ذو المتغيرين – شكل الانتشار – تقدير البارامترين α ، β طريقة المربعات الصغرى – الخطأ المعيارى لخط الانحدار – استنتاجات إحصائية – التوسع فى استخدام الانحدار الخطى البسيط – معنى آخر للانحدار – الاختلاف المفسر وغير المفسر – معامل التحديد – تحليل الانحدار حين وجود أكثر من قيمة ص لكل قيمة س – اختبار جودة العلاقة الخطية – ملاحظات عن افتراضات الانحدار – استخدامات الانحدار .

الفصل العاشر : الارتباط الخطى البسيط .... ٤١٣ – ٤٣٩
الانحدار والارتباط – افتراضات الارتباط الخطى البسيط – معامل الارتباط العزمى (بيرسون) – مميزات معامل الارتباط العزمى – دلالة معامل الارتباط العزمى – اختبار $\rho = 0$ – اختبار ρ = ك – اختبار دلالة الفرق بين معاملى ارتباط عينتين مستقلتين – التمييز بين الانحدار والارتباط فى دراسة

المشكلات – معامل ارتباط الرتب (سبيرمان) – مميزات معامل ارتباط الرتب – دلالة معامل ارتباط الرتب .

الفصل الحادى عشر : تحليل التغاير .... ٤٤١ – ٤٦٨

التغاير – العلاقة بين تحليل التغاير وتحليل التباين – النموذج الإحصائى – خطوات تحليل التغاير – المقارنة بين المتوسطات المعدلة – تحليل التغاير فى التجارب ذوات العاملين – المقارنة بين أزواج المتوسطات .

الفصل الثانى عشر : الانحدار والارتباط الخطى المتعدد .... ٤٦٩ – ٥١٠

(أولا) **الانحدار الخطى المتعدد** – كامتداد للانحدار الخطى البسيط – إيجاد معادلة الانحدار – إيجاد الخطأ المعيارى للتقدير من خط الانحدار – اختبار دلالة الانحدار ككل –اختبار دلالة معاملات الانحدار الجزئية – استخدام الحاسب الالكترونى – طريقة دوليتل لحل المعادلات الخطية – أسلوب آخر لاختبار دلالة معاملات الانحدار الجزئية – اختيار متغيرات التنبؤ .

(ثانيا) **الارتباط الخطى المتعدد** : معامل الارتباط الخطى المتعدد واختبار دلالته – معامل الارتباط الجزئى واختبار دلالته .

الفصل الثالث عشر : دالة التمييز .... ٥١١ – ٥٢٨

دالة التمييز – إيجاد دالة التمييز (حالة متغير واحد وحالة ك من المتغيرات) –افتراضات التحليل–النقطة الحدية – احتمال خطأ التقسيم – اختبار تساوى أزواج المتوسطات – اختبار ت^2 – استخدامات دالة التمييز .

الفصل الرابع عشر : الطرق غير البارامترية .... ٥٢٩ – ٥٦٣

اختبار التلاحقات – اختبار الإشارة – اختبار ويلكوكسن للمقارنات التزاوجية – اختبار الاتجاه – اختبار كروسكال – واليس ، اختبار فريدمان

الفصل الخامس عشر : اختيار العينات وتحليلها .... ٥٦٥ – ٦٠٦

المعاينة العشوائية – المعاينة الاحتمالية – العينة العشوائية البسيطة – العينة الطبقية (طريقة التقسيم المتناسب – طريقة التقسيم الأمثل – تقدير البارامترات والأخطاء المعيارية – المعاينة الطبقية من مجتمع ذى حدين) – العينة المتعددة المراحل (التحليل الاحصائى – الاختبارات الإحصائية) – العينة المنتظمة – المعاينة المساحية – العينات غير الاحتمالية .

الملحق (١) أجوبة التمارين .... ٦٠٧ – ٦١٧

الملحق (٢) جداول إحصائية .... ٦١٩ – ٦٤١

مراجع ٦٤٣

الفصل الأول

مفاهيم أولية

يتناول هذا الفصل عدداً من المفاهيم التي ترتكز عليها دراسة الإحصاء التطبيقي خاصة في مجال العلوم ، فهي بمثابة الأبجدية التي لا مفر من إرسائها قبل متابعة تلك الدراسة . وهي وإن بدت منفصلة عن بعضها إلا أنها تتكاتف معا لخدمة موضوعات الفصول التالية .

(١ - ١) المجتمع والعينة :

إن التمييز بين المجتمع والعينة هو أول ما ينبغي أن يتنبه له أي باحث تطبيقي خاصة حين يستخدم الطرق الإحصائية وعملية الاستدلال الإحصائي .

في الإحصاء تستخدم كلمة **مجتمع** للتعبير عن أي مجموعة (منتهية أو لا نهائية) من الأشياء أو الأحداث التي تكون موضع اهتمامنا في وقت ما من حيث ظاهرة ما أو متغير ما س . ومن أمثلة المجتمعات الإحصائية : جميع نباتات حنك السبع المزروعة في حقل ما في وقت ما ، جميع نباتات البسلة في العالم ، جميع القواقع في بحيرة ما ، جميع الفيران من نوع معين ، سقوط المطر في منطقة ما ...

وينبغي أن يكون المجتمع الذي ندرسه معرفا تعريفا جيدا خاصة فيما يتعلق بالمتغير س وطريقة قياسه وفي تحديد الوحدة أو العنصر التي يتكون المجتمع من مجموعها . فمثلا قد يعبر المتغير س عن الطول أو الوزن أو اللون أو عدد ضربات القلب التي تنتج عن حقن قلب الفأر بمادة الأدرينالين ... وقد تكون الوحدة هي قرن بسلة أو قوقعة أو قلب فأر ...

وفي كثير من الأحيان يمكن وصف المتغير س في مجتمع ما عن طريق نموذج نظرى يوضع على هيئة معادلة أو صيغة رياضية تعبر عما يسمى « توزيع المجتمع » فنقول مثلا إن مجتمعا ما هو مجتمع معتدل أو مجتمع ذو حدين أو مجتمع بواسوني ... وهذا ما سوف نتناوله فيما بعد .

والأعداد التي تميز مجتمعا ما تسمى بارامترات (أو معالم أو أدلة أو ثوابت) المجتمع وهى أعداد ثابتة تميز كل مجتمع عن غيره من المجتمعات حتى ولو كان له نفس التوزيع ، مثل الوسط الحسابي μ وهو مقياس للنزعة المركزية للمجتمع ، والانحراف المعيارى σ وهو مقياس لتشتت مفردات المجتمع حول الوسط الحسابى .

أما **العينة** فهي جزء من المجتمع يختار بحسب مواصفات معينة وبهدف استخدامها لدراسة المجتمع ، وهناك من النظريات والطرق الإحصائية ما يمكننا من تقدير بارامترات المجتمعات أو مقارنتها أو إصدار قرارات أو تنبؤات بشأنها عن طريق فحص ودراسة عينات مأخوذة منها .

وبطبيعة الحال ينبغي أن تختار العينة بحيث تمثل المجتمع أفضل تمثيل ممكن ، على أن التحليل الإحصائي يتطلب بالضرورة أن تكون العينة **عشوائية** . والعشوائية لاتعني أن نأخذ جزءا من المجتمع بشكل جزافي ، بل هي إجراء يصمم بدقة بحيث يضمن عدم وجود تحيز من أى نوع قد يؤثر في عملية اختيار العينة وبحيث يكون لكل وحدة من وحدات المجتمع احتمال معروف للدخول في العينة . ومن ثم فقد وضعت خطط مختلفة للمعاينة العشوائية **random sampling plans** أى لسحب العينات العشوائية ، يتوقف استخدام أى منها لدراسة مجتمع ما على طبيعة هذا المجتمع وعلى الهدف من دراسته . ومن أشهر هذه الخطط ما ينتج العينات التي تحمل الأسماء الآتية : العينة العشوائية البسيطة – العينة الطبقية – العينة ذات المراحل – العينة المنتظمة – العينة المساحية ... وسنهتم هنا بصفة خاصة بالعينة العشوائية البسيطة التى هى على أية حال أساس لكثير من تلك الخطط ، أما بقية الخطط فقد أفرد لها فصل خاص هو الفصل الخامس عشر من هذا الكتاب .

(١ – ٢) العينة العشوائية البسيطة : SIMPLE RANDOM SAMPLE

لعل أبسط تعريف للعينة العشوائية البسيطة هو أنها تلك العينة التى تؤخذ من المجتمع بحيث يكون لكل وحدة من وحداته نفس الفرصة في الظهور في العينة . ولذلك فإن هذه العينة لا تصلح لتمثيل المجتمع إلا إذا كان هذا المجتمع **متجانسا** من حيث المتغير الذي ندرسه . وفيما يلي حين نذكر كلمة عينة فسوف نقصد عينة عشوائية بسيطة .

ولعل أوضح خطة لاختيار عينة عشوائية بسيطة من مجتمع منتهي هي تلك التى اشتهر استخدامها في تحديد أصحاب جوائز المسابقات التليفزيونية . نفرض مثلا أن لدينا ٤٥٢ شخصا أجابوا إجابات صحيحة ونريد أن نختار منهم ٢٠ شخصا دون تحيز . إن هذا يعني أن لدينا مجتمعا متجانسا حجمه ٤٥٢ ونريد سحب عينة عشوائية بسيطة حجمها ٢٠ . نعد ٤٥٢ قطعة صغيرة متطابقة من الورق ونكتب عليها أسماء هؤلاء الأشخاص (أو نعطيهم أرقاما مسلسلة من ١ إلى ٤٥٢ ثم نكتب هذه الأرقام على قطع الورق) . نضع هذه القطع في علبة ونخلطها جيدا ثم نسحب منها ٢٠ قطعة الواحدة بعد الأخرى دون أى تحيز مع إرجاع كل قطعة تسحب إلى الصندوق قبل كل سحبة وخلط القطع جيدا في كل مرة وإهمال القطع التى تتكرر في السحب ، فتكون مجموعة الأشخاص الذين تظهر أسماؤهم أو أرقامهم على هذه القطع هى العينة المطلوبة .

غير أنه يمكن الاستغناء عن العلبة وقطع الورق وبناء خطة الاختيار على أحد الجداول المسماة **بجداول الأرقام العشوائية** ، مثل الجدول (١) بملحق هذا الكتاب . ويتألف هذا الجدول من عدة أعمدة (أو صفوف) مجمعة كل خمسة معا للسهولة وبكل عمود أرقام ٠ ، ١ ، ٢ ، ... ، ٩ مرتبة ترتيبا خاصا يجعل لكل من هذه الأرقام نفس الفرصة في الظهور في أي موضع بالجدول ، ويمكن أن نقرأ منها أعدادا متتابعة يتألف كل منها من رقم واحد أو رقمين أو ثلاثة .. أى نأخذ عمودا واحدا أو اثنين أو ثلاثة ... حسبما نريد ، كما يمكن أن نبدأ القراءة

من أى صف أو أى عمود وفي أى اتجاه ، إلى أعلى أو إلى أسفل يمينا أو يسارا أو قطريا . ويعتمد عدد الأعمدة التي نختارها على عدد الأرقام التي يشتمل عليها حجم المجتمع .

مثال (١ – ١) :

لدينا مجتمع حجمه ٥٠٠ من المرضى بمرض معين ونريد أن نستخدم جدول الأرقام العشوائية لاختيار عينة عشوائية بسيطة حجمها ٢٠ لإجراء بعض القياسات على عناصرها .

نظرا لأن حجم المجتمع وهو ٥٠٠ يتألف من ثلاثة أرقام ، نعطى لوحدات المجتمع أرقاما مسلسلة ٠٠١ ، ٠٠٢ ، ... ، ٠١٠ ، ٠١١ ، ... ، ٤٩٩ ، ٥٠٠ يتألف كل منها من ٣ خانات ومن ثم نحتاج لاختيار العينة إلى ثلاثة أعمدة . نغمض أعيننا ونضع أصبعنا عشوائيا على أى نقطة في الجدول ولتكن هذه النقطة هى نقطة التقاء العمود ٨ والصف ٢١ حيث نجد العدد ٠٠٨ فيكون هذا الرقم هو الرقم المسلسل الأول في العينة . نقرأ ابتداء من هذا العدد رأسيا إلى أسفل (مثلا) بشكل هندسي ثابت مع إهمال الأعداد التي تزيد عن ٥٠٠ والأعداد التي يتكرر ظهورها . وإذا صادف أن انتهى العمود ٨ ولم يكتمل بعد الحجم المطلوب للعينة نتحول إلى ثلاثة أعمدة أخرى من أى من الجانبين الأيمن أو الأيسر ونقرأ من أعلى الجدول أو من أى مكان آخر في نفس الاتجاه السابق حتى نصل إلى الحجم المطلوب ونكون بذلك قد توصلنا إلى عينة عشوائية بسيطة قد تأخذ مفرداتها الأرقام المسلسلة الآتية :

٠٠٨ ١٤٢ ٠١٨ ١٤٣ ٤١٨ ٤٩١ ٢٤٠ ٠٢٠ ١٤٠ ٠٠٦
١٠١ ٢٤٩ ٠١٦ ١٣٤ ٠٤٩ ٣١٩ ٤٨٣ ٢٣٢ ٠٥٢ ٠٢٩

ملاحظة : احتمال الحصول على عينة بهذه الطريقة يساوى احتمال الحصول على أى عينة أخرى من نفس الحجم ، وتؤخذ هذه الخاصة أحيانا كتعريف للعينة العشوائية البسيطة

(١ – ٣) المتغيرات الإحصائية : STATISTICAL VARIABLES

في أي دراسة تطبيقية إحصائية ينبغي أن نحصل على بيانات عددية عن المتغيرات التي ندرسها . وتختلف طريقة تناولنا لهذه البيانات باختلاف تلك المتغيرات التي يمكن تقسيمها بصفة عامة إلى الأنواع الآتية :

أولا – المتغيرات الوصفية (أو النوعية) :

QUALITATIVE VARIABLES (ATTRIBUTES)

وهى متغيرات لا يمكن قياس مفرداتها عدديا ولكن يمكن تقسيم هذه المفردات بحسب اشتراكها في صفة أو خاصة ما . ومن أمثلة هذه المتغيرات متغير اللون أو الجنس أو المهنة أو الخواص الوراثية . والجدول (١ – ١) هو مثال لبيانات وصفية فهو يعطى التوزيع التكرارى لألوان عيون عينة من ٨٠ فارا .

الجدول (١ – ١)
التوزيع التكرارى لألوان عيون عينة من الفيران

اللون	التكرار
أسود	٤
ياقوتي	٢
رمادي	٧٤
المجموع	٨٠

الجدول (١ – ٢)

ترتيب محكمين لخمسة من الأشخاص

المتقدم	المحكم الأول	المحكم الثاني
أ	٢	١
ب	١	٢
جـ	٤	٥
د	٣	٤
هـ	٥	٣

ثانيا – المتغيرات الرتبية : RANKED VARIABLES

وهى أيضا متغيرات لا يمكن قياس مفرداتها عدديا ولكن يمكن تنظيم هذه المفردات بحسب ترتيب ما أو رتبة ما . ومن أمثلة هذه المتغيرات متغير النمو الذى

يمكن تقسيمه إلى ضعيف – عادى – مفرط ، كذلك المتغيرات التي تنتج من عملية التحكيم كما هو الحال مثلا عندما يطلب إلى مجموعة من علماء النبات ترتيب عشرة نباتات من حيث شدة التلف الذى أصابها من مرض فطرى ، فيعطى كل منهم بحسب تقديره الترتيب (١) لأقل النباتات تأثرا بالمرض والترتيب (١٠) لأكثرها تأثرا ، أو حينما يطلب إلى مجموعة من الأطباء إبداء آرائهم في مشاهداتهم الميكروسكوبية عن مرض السرطان وترتيبها من حيث تزايد الورم السرطانى . الجدول (١ – ٢) يعطى الترتيب الذى رآه اثنان من المحكمين لخمسة من المتقدمين لشغل وظيفة ما .

ثالثا – المتغيرات العددية « الكمية » : QUANTITATIVE VARIABLES

وهى التى يمكن التعبير عنها بالأعداد العادية (الحقيقية) فيكون التغير فيها هو تغير من حيث المقدار أى يمكن تقسيم مفرداتها بحسب الأصغر والأكبر ، ونميز هنا بين النوعين الآتيين :

(ا) المتغيرات المتصلة : CONTINUOUS VARIABLES

وهى تلك التى نحصل عليها عادة عن طريق القياس measurement مثل الطول والوزن والعمر ودرجة الحرارة . وفى هذه المتغيرات نستطيع أن نتصور أن المفردات يمكن أن تختلف عن بعضها بمقادير صغيرة صغراً لا نهائيا من الناحية النظرية . وإذا وقعت قيم متغير متصل بين عددين ا ، ب فإنها تمثل بيانيا بجميع نقط قطعة مستقيمة محدودة بهذين العددين .

(ب) المتغيرات الوثابة : DISCRETE (OR MERISTIC) VARIABLES

وهى تلك التى نحصل عليها عادة عن طريق العد counting مثل عدد الذرية – عدد الخلفة – عدد ضربات القلب . ومجموعة قيم المتغير الوثاب قد تكون منتهية مثل { ٠ ، ١ ، ٢ ، ... ، ن } أو لا نهائية مثل {٠ ، ١ ، ٢ ، ...} . وفى كلتا الحالتين تمثل بيانيا بنقط منفصلة على خط مستقيم .

وفى المتغيرات العددية نميز من جهة أخرى بين المتغيرات **العشوائية** والمتغيرات **غير العشوائية** . والمتغيرات العشوائية هى متغيرات تخضع لمؤثرات عشوائية لا يمكن التحكم فيها تجريبيا ومن أمثلتها متغير درجات حرارة الجو ومتغير ضغط دم الإنسان أو تحصيله الدراسى ومتغير عدد ضربات قلب ضفدعة بعد معالجة ما ، وهذه جميعها تتأثر بعوامل عشوائية غير منظورة . أما المتغيرات غير العشوائية (أو الرياضية) فهى تلك التى لا تكون واقعة تحت تأثير عوامل عشوائية ولذلك يمكن قياسها بدقة أو بأخطاء صغيرة يمكن إهمالها .

ويهتم الإحصاء بصفة خاصة بالمتغيرات العشوائية بل هى محور الدراسة فيه نظريا وتطبيقيا ولذلك سوف نتناولها بشيء من التفصيل هى وتوزيعاتها ونماذجها في الفصل الثالث من هذا الكتاب ليتيسر لنا التعامل معها بعد ذلك في الفصول التالية .

(١ – ٤) الأخطاء العشوائية (أخطاء الصدفة) RANDOM ERRORS

هناك أخطاء تخص المجال الذى يعمل فيه الباحث ولا تدخل في صميم الموضوعات الإحصائية كالأخطاء الناتجة عن عدم ضبط الأجهزة أو الأدوات المستخدمة في القياس أو عدم توفر الظروف الملائمة لإجراء التجربة كدرجة الحرارة أو الرطوبة ، أو الأخطاء الناتجة عن عدم ملاءمة طريقة القياس ، أى عندما يكون هناك اختلاف بين التعريف النظرى للمتغير والتعريف الإجرائى المستخدم فى عملية قياس هذا المتغير .

أما ما نهتم به فى الإحصاء فهى الأخطاء العشوائية ، وتظهر فى الاختلافات التى نجدها فى القيم العددية التي نحصل عليها عند القيام بقياسات متكررة لنفس الوحدة أو الشيء ، كما هو الحال مثلا عند قياس طول حشرة عدة مرات . ويتسبب فى هذه الاختلافات عدد كبير من العوامل الثانوية التي نعجز عن حصرها أو تحديد مصدرها أو التنبؤ بها ونعجز عن حساب التأثير الضئيل الذى يحدثه كل منها على حدة . وبالتالى نعجز عن التحكم فيها تجريبيا ، وفى هذه الحال نحاول البحث عن وسائل تسمح بتقدير تأثيرها الكلى للإفادة من هذا التقدير فى عملية التحليل الإحصائى

وحين تكون القياسات معرضة للأخطاء العشوائية فقط فإن أى قيمة $س_ر$ نحصل عليها من قياس عنصر منه تمثل بالنموذج الآتى :

$$س_ر = أ + خ_ر .$$

حيث أ هى القيمة الحقيقية للعنصر المقاس ، $خ_ر$ هى الخطأ فى $س_ر$ أى انحراف $س_ر$ عن القيمة الحقيقية أ .

إلا أنه على المدى البعيد تميل هذه الأخطاء إلى تعويض بعضها البعض بمعنى حدوث توازن بين الأخطاء الموجبة والأخطاء السالبة بحيث يقترب متوسط القياسات من القيمة الحقيقية للشىء الذى نقيسه ، ولذلك فإننا نفترض فى كثير من الحالات أن متوسط الأخطاء هو صفر على المدى البعيد .

(١ - ٥) قواعد تقريب الأعداد : RULES FOR ROUNDING

كثيرا ما نلجأ إلى تدوين قياساتنا للمتغيرات العددية مقربة إلى عدد معين من الخانات . ولقد وجد أنه من المناسب الاتفاق على القواعد الآتية :

ليكن ق هو الرقم الذى فى الخانة المراد التقريب إليها ، أ هو الرقم التالى من اليمين مباشرة للرقم ق .

(أولا) إذا كان أ < ٥ يبقى الرقم ق كما هو .

(ثانيا) إذا كان أ > ٥ يزاد الرقم ق واحدا .

(ثالثا) إذا كان أ = ٥ ، أ متبوعا بأرقام غير الصفر ، يزاد الرقم ق واحدا .

أما إذا كان أ = ٥ ، أ يقف وحده أو متبوعا بأصفار ، يبقى الرقم ق كما هو إذا كان زوجيا ويزاد واحدا إذا كان فرديا . وهذه القاعدة الأخيرة من شأنها إحداث توازن بين الأعداد التى زيدت فى التقريب والأعداد التى نقصت ، خاصة إذا كان لدينا متتابعة طويلة من الأعداد المقربة .

مثال (١ – ٢)

العدد أ ق	العدد مقربا (إلى خانتين عشريتين)
٤٨,٣٢٤٦	٤٨,٣٢
٤٨,٣٢٦٠	٤٨,٣٣
٤٨,٣٢٥١	٤٨,٣٣
٤٨,٣٢٥٠	٤٨,٣٢
٤٨,٣١٥٠	٤٨,٣٢

مثال (١ – ٣) :

العدد	عدد الأرقام المعنوية المطلوبة	العدد المقرب
٢٦,٥٨	٢	٢٧
١٣٣,٧١٤٧	٥	١٣٣,٧١
١٣٣,٧١٤٧	٦	١٣٣,٧١٥
,٠٣٧٢٥	٣	,٠٣٧٢
,٠٣٧١٥	٣	,٠٣٧٢
١٨٣١٦	٢	١٨٠٠٠
١٨٣١٦	٣	١٨٣٠٠
١٧,٣٤٧٦	٣	١٧,٣

هذا مع ملاحظة أن آخر رقم فى عدد مقرب ينبغى دائما أن يكون معنويا أى ينبغى أن يتضمن فترة تقع فيها القيمة الحقيقيةللعدد ، وهذهالفترة تبدأ بنصف وحدة خطوة أسفل العدد المقرب المدون وتنتهى بنصف وحدة خطوة أعلاه فمثلا العدد المقرب ٧,٨ يعنى أن قيمته الحقيقية تقع في الفترة بين ٧,٧٥ ، ٧,٨٥ أى ٧,٨ ± ٠,٠٥ والعدد المقرب ٧,٨٠ يعنى أن قيمته الحقيقية تقع في الفترة بين

٧,٧٩٥ ، ٧,٨٠٥ أى٧,٨٠ ± ٠,٠٠٥ والعدد المقرب ٣,٢٤ يعنى أن قيمته الحقيقية تقع فى الفترة بين ٣,٢٣٥ ، ٣,٢٤٥ أى ٣,٢٤ ± ٠,٠٠٥

وحين نرغب فى تدوين قياس وحدة ما مقربة إلى عدد معين من الخانات نوجد هذا القياس بحيث يكون عدد الخانات الناتجة أكثر بواحد على الأقل من العدد المطلوب ثم نجرى التقريب . فمثلا لإيجاد خارج القسمة ١٥ ÷ ٧ مقربا إلى ٣ خانات عشرية نقوم بعملية القسمة حتى نحصل على ٤ خانات ثم نقرب إلى ٣ خانات كالآتى : ١٥ ÷ ٧ = ٢,١٤٢٨ = ٢,١٤٣ تقريبا .

(١ – ٦) الدقة والضبط : ACCURACY AND PRECISION

الدقة هى تعبير عن مدى قرب قيمة نتجت عن قياس وحدة ما من القيمة الحقيقية لهذه الوحدة ، أما الضبط فهو تعبير عن مدى قرب القياسات المتكررة لوحدة ما من بعضها البعض تحت نفس الظروف . والإحصاء يهتم أساسا بالضبط لأن الضبط يتضمن الدقة طالما كانت أداة القياس غير متحيزة .

والدقة في عدد مقرب يحكم عليها بدلالة النسبة المئوية للخطأ الذى يحتويه ، فمثلا نفرض أن عددا سجل على أنه ٨٩ تقريبا . إن هذا يعنى أن القيمة الحقيقية لهذا العدد تقع بين العددين ٨٨,٥ ، ٨٩,٥ وتكون القيمة المطلقة للحد الأقصى للخطأ هى ٠,٥ وبالتالى تكون نسبة الخطأ هى :

$$\frac{٠,٥}{٨٩} \times ١٠٠ = ٠,٥٦ .$$

نفرض أن عددا آخر سجل على أنه ١٥,٥ تقريبا فتكون نسبة الخطأ هى :

$$\frac{٠,٠٥}{١٥,٥} \times ١٠٠ = ٠,٣٢$$

ونظرا لأن ٠,٥٦ أكبر من ٠,٣٢ نقول إن العدد ٨٩ أقل دقة من العدد ١٥,٥ أى أن العدد يكون أكثر دقة إذا استطعنا كتابته بعدد أكبر من الأرقام المعنوية .

أما الضبط فنحكم عليه بعدة طرق منها حجم الوحدة المستخدمة في القياس ، ففى قياس طول وحدة ما يكون القياس أكثر ضبطا حين نستخدم مسطرة مدرجة بالمليمترات عنه حين نستخدم مسطرة مدرجة بالبوصات . وفى الإحصاء يتم التعبير عن الضبط فى كثير من الحالات بواسطة الانحراف المعيارى للقياسات المتكررة .

(١ – ٧) الاحتمال : PROBABILITY

إن القرارات والأحكام والتنبؤات التى نتخذها فى الإحصاء هى دائما قرارات احتمالية بمعنى أننا لا نستطيع أن نجزم جزما باتا بصحتها أو بخطئها ، فتأتى القرارات مصحوبة باحتمالات معينة للصواب أو الخطأ . ولعل هذا هو الذى أعطى الإحصاء قوته الكبرى وميزته عن الرياضيات ، إذ يتعرض للقضايا والفروض التى تعلن الرياضيات عجزها عن تناولها ، فيبت فيها بأسلوب احتمالى هو على أية حال أفضل من ترك القضايا دون حل عملا بالحكمة القائلة : « ما لا يدرك كله لا يترك كله ». ويهمنا فى مستهل دراستنا للإحصاء أن نتبين معنى الاحتمال من الناحية التطبيقية على الأقل .

إذا كان لدينا فرض ما فإننا نعين لصواب هذا الفرض العدد «١» ونعين لخطئه العدد « صفر » ومن ثم فإن أى عدد يقع بين الصفر والواحد يمكن أن يعبر عن درجة صواب أو خطأ هذا الفرض . إن الأعداد التى تبدأ بالصفر وتنتهى بالواحد هى التى نعبر بها عن الاحتمالات ، فنقول مثلا إن احتمال ظهور البترول فى منطقة ما هو ٠,٦ أو إن احتمال الحصول على نباتات حمراء من مجموعة معينة من بذور حنك السبع هو ٠,٤ وهكذا ... ويؤخذ الاحتمال بداهة كمقياس لدرجة اعتقادنا أو شدة اقتناعنا فى صحة فرض ما أو فى وقوع حدث ما . وليس من المناسب أن نعتبر هذا تعريفا للاحتمال لأنه يخضع لذاتية المشاهد وخبرته ، ومع ذلك فكثيرا ما نجد أنفسنا مضطرين إلى أن نقدر احتمال حدث ما بالبداهة أو الخبرة أو أى عوامل ذاتية أخرى وذلك حين لا تتوفر عوامل مباشرة تكفى لحساب احتمال الحدث .

والتعريف الدقيق للاحتمال يتألف من مجموعة من المسلمات المصاغة صياغة رياضية ، على أن الدراسات التطبيقية لا تحتاج إلى هذا التعريف وإنما تكتفى بتعريفين فى مستوى أقل ، فيعتبر الأول أن الاحتمال هو « نسبة » ويعتبر الثانى أن الاحتمال هو « تكرار نسبى » كما هو موضح بعد .

(أ) التعريف التقليدى : CLASSICAL DEFINITION

إذا أسفرت تجربة عن ن من النواتج أو الحالات المتساوية الإمكانات وكان الحدث أ يقع فى م من هذه النواتج فإن احتمال وقوع هذا الحدث يساوى $\frac{م}{ن}$

أى يساوى النسبة بين عدد الحالات التى يمكن أن يقع فيها الحدث والعدد الكلى للحالات التى يمكن أن تسفر عنها التجربة .

فمثلا إذا ألقينا حجرة نرد منتظمةعشوائيا تكون النواتج الممكنة للتجربة هى الأعداد الست ١ ، ٢ ، ٣ ، ٤ ، ٥ ، ٦ وهذه النواتج متساوية الإمكانات لأنه لا يوجد لدينا ما يجعلنا نتوقع ظهور أحدها دون الآخرين وعلى ذلك فإن احتمال الحدث « عدد فردى » مثلا هو $\frac{٣}{٦} = \frac{١}{٢}$. كذلك احتمال سحب « صورة » من مجموعة محكمة الخلط من ورق اللعب = $\frac{١٢}{٥٢} = \frac{٣}{١٣}$.

(ب) التعريف الإحصائى : STATISTICAL DEFINITION

ينطلق هذا التعريف من فكرة التكرار النسبى ومن ظاهرة اكتشفت بالملاحظة والتجريب تعرف **بظاهرة الانتظام الإحصائى Statistical regularity** ومجملها أنه إذا كررت تجربة مرات كثيرة تحت نفس الظروف (مثل زراعة بذرة من نبات حنك السبع) فإن التكرار النسبى لحدث ما متعلق بهذه التجربة (مثل ظهور اللون الأحمر) يقترب من عدد ثابت كلما زاد عدد مرات إجراء التجربة ، ويؤخذ هذا العدد كتقدير لاحتمال ذلك الحدث . ولذلك يعرف احتمال حدث ما بأنه نهاية متتابعة من التكرارات النسبية لوقوع هذا الحدث . فمثلا إذا ألقينا قطعة نقود منتظمة عشوائيا عشر مرات ثم مائة ثم ألف ثم ... فإن التكرار النسبى لظهور

الصورة يقترب من العدد $\frac{1}{2}$ كلما زاد عدد مرات إلقاء القطعة . وهنا نقول إن احتمال الحدث « ظهور الصورة » هو $\frac{1}{2}$. ويفهم من هذا أنه إذا رميت قطعة نقود عددا كبيرا من المرات فإن نسبة عدد المرات التى تظهر فيها الصورة إلى العدد الكلى لمرات رمى القطعة هو $\frac{1}{2}$ (على المدى الطويل) .

ويوحى التعريف الإحصائى بطريقة مناسبة لإيجاد احتمالات الأحداث تجريبيا ، فإذا أردنا مثلا إيجاد احتمال وجود وحدة معيبة من الوحدات التى ينتجها مصنع ما ، نسحب عينة عشوائية كبيرة من هذه الوحدات ونحسب التكرار النسبى للوحدات المعيبة فنحصل على تقدير للاحتمال المطلوب ، ويمكن بعد ذلك اختبار دقة هذا التقدير بالطرق الإحصائية التى سندرسها بعد .

أما التعريف التقليدى فيصلح لإيجاد الاحتمالات نظريا فى الحالات التى يتوفر فيها شرط تساوى الإمكانات كما فى المثال الآتى .

مثال (١ - ٤) :

فى مجموعة من ١٠٠ رجل علم أن ١٢ منهم يلبسون نظارات ، ٨ منهم لا يلبسون نظارات ولكن يحتاجون إليها . إذا اخترنا رجل واحد من هذه المجموعة عشوائيا فإن :

(ا) احتمال أن يكون الرجل لابساً نظارة = ٠,١٢

(ب) احتمال أن يكون الرجل « لا يلبس نظارة ولكن يحتاج إليها » = ٠,٠٨

(ج) احتمال أن يكون الرجل « لا يلبس نظارة ولا يحتاج إليها » = ٠,٨٠

امتداد للتعريف التقليدى :

إن التعريف التقليدى للاحتمال الذى سبق تقديمه يتناول التجارب التى تسفر عن عدد منتهى من النواتج ، إلا أنه مع بعض التعديل يمكن أن يمتد ليشمل التجارب التى تسفر عن عدد غير منتهى من النواتج كتلك التى يكون فيها لفضاء التجريب مقياس هندسى كالطول أو المساحة أو الحجم . فإذا كان هذا الفضاء يتألف من

منطقة محددة ا وكانت المنطقة ب جزءا من المنطقة ا واخترنا عشوائيا نقطة من المنطقة ا فإن احتمال الحدث « وقوع النقطة فى المنطقة ب » يعرف بالنسبة :

$$\frac{\text{مقياس ب}}{\text{مقياس ا}}$$

وذلك مع الاحتفاظ بفرض تساوى الإمكانات . ويلاحظ هنا أن تعبير العشوائية يعنى أن احتمال وقوع نقطة فى جزء من فضاء التجريب يتناسب مع مقياس هذا الجزء وأن هذا الاحتمال مستقل عن شكل المنطقة وموضعها .

مثال (١ – ٥)

اختيرت نقطة عشوائيا من داخل دائرة ا نصف قطرها ٣ سم . ما احتمال ألا يزيد بعد هذه النقطة عن مركز الدائرة عن ٢ سم ؟

الحل :

يقع الحدث المطلوب إذا وقعت النقطة المختارة داخل دائرة ب لها نفس مركز الدائرة ا ونصف قطرها ٢ سم .

$$\therefore \text{الاحتمال المطلوب} = \frac{\text{مقياس الدائرة ب}}{\text{مقياس الدائرة ا}} = \frac{ط \times ٤}{ط \times ٩} = \frac{٤}{٩}$$

مثال (١ – ٦)

اختيرت نقطة عشوائيا على القطعة المستقيمة ا = { س : ١٠ ≥ س ≥ ٠ }
ما احتمال وقوع النقطة فى القطعة المستقيمة ب = { س : ٨ ≥ س ≥ ٢ }

الحل :

$$\text{الاحتمال المطلوب} = \frac{\text{طول القطعة ب}}{\text{طول القطعة ا}} = \frac{٦}{١٠} = ٠{,}٦$$

اصطلاحات وتعاريف :

(١) الرمز ل (ا) يعنى احتمال وقوع الحدث ا ، ويلاحظ أن

١ ⩾ ل (ا) ⩾ ٠

وحين ل (ا) = ٠ نقول إن الحدث ا هو حدث مستحيل ،

وحين ل (ا) = ١ نقول إن الحدث ا هو حدث مؤكد .

(٢) الرمز ل (ا$_1$ أو ا$_2$)

يعنى احتمال وقوع واحد على الأقل من الحدثين ا$_1$ ، ا$_2$ أى وقوع ا$_1$ فقط أو ا$_2$ فقط أو ا$_1$ ، ا$_2$ معا .

(٣) الرمز ل (ا$_1$ وا$_2$)

يعنى احتمال وقوع الحدثين ا$_1$ ، ا$_2$ معا أو بالتتابع .

(٤) الرمز ل (ا$_2$ | ا$_1$)

يعنى احتمال وقوع الحدث ا$_2$ بشرط أن يكون الحدث ا$_1$ قد وقع فعلا ويسمى هذا الاحتمال بالاحتمال الشرطى للحدث ا$_2$ بالنسبة للحدث ا$_1$.

مثال ذلك احتمال اختيار طالب من مدرسة ما بشرط أن يكون من الرياضيين .

(٥) يقال للحدثين ا$_1$ ، ا$_2$ إنهما **متنافيان** إذا كان وقوع أحدهما يمنع وقوع الآخر وبذلك لا يمكن أن يقع الحدثان معا ، أى يكون ل (ا$_1$ وا$_2$) = ٠ فمثلا فى حالة ولادة طفل يكون الحدث « المولود ذكر » والحدث « المولود أنثى » حدثين متنافيين (لا يمكن أن يقعا معا) .

(٦) يقال للحدثين ا$_1$ ، ا$_2$ إنهما **مستقلان** إذا كان احتمال وقوع أيهما لا يتأثر بوقوع أو عدم وقوع الآخر ، أى إذا كان :

$$ل(أ_٢ | أ_١) = ل(أ_٢)، \quad ل(أ_١ | أ_٢) = ل(أ_١)$$

فمثلا إذا ألقينا قطعتين من العملة عشوائيا فإن ما يظهر على أى منهما (صورة أو كتابة) يكون مستقلا عما يظهر على الأخرى (إلا إذا كانت القطعتان مربوطتين معا بخيط مثلا) . كذلك اختيار طالب من كلية العلوم واختيار طالب من كلية الإنسانيات هما حدثان مستقلان .

توافقات الاحتمال :

القاعدتان الآتيتان تسهلان حساب الاحتمالات في كثير من الحالات ويمكن إثباتهما رياضياً :

أولا – قاعدة الجمع :

$$ل(أ_١ أو أ_٢) = ل(أ_١) + ل(أ_٢) - ل(أ_١ و أ_٢)$$

أى أن احتمال وقوع أحد الحدثين $أ_١$ ، $أ_٢$ أو كلاهما يساوى

احتمال وقوع الأول + احتمال وقوع الثانى – احتمال وقوعهما معا .

حالة خاصة :

إذا كان الحدثان $أ_١$ و $أ_٢$ **متنافيين** فإنه حسب التعريف (٥) تصبح قاعدة الجمع كالآتى :

$$ل(أ_١ أو أ_٢) = ل(أ_١) + ل(أ_٢)$$

وتسمى هذه القاعدة **بقاعدة الجمع للأحداث المتنافية** ، ويمكن تعميمها كالآتى :

إذا كانت $أ_١$ ، $أ_٢$ ، ... ، $أ_ن$ أحداثا متنافية فإن :

$$ل(أ_١ أو أ_٢ أو ... أو أ_ن) = ل(أ_١) + ل(أ_٢) + ... + ل(أ_ن)$$

مثال (١ – ٧)

فى مجموعة من ١٠٠ طالب رسب ١٢ فى الرياضيات ورسب ١٥ فى الفيزياء ورسب ٨ فى كلتا المادتين . إذا اختير طالب واحد عشوائيا من هذه المجموعة فما احتمال أن يكون راسبا فى الرياضيات أو فى الفيزياء ؟

الحل :

نلاحظ أن الرسوب فى أى مادة لا يمنع الرسوب فى المادة الأخرى ، أى أن الحدثين غير متنافيين وحسب قاعدة الجمع نجد أن :

ل (راسب فى الرياضيات أو راسب فى الفيزياء) = ٠,١٢ + ٠,١٥ – ٠,٠٨ = ٠,١٩

مثال (١ – ٨) :

ألقى حجر نرد منتظم عشوائيا مرة واحدة . نجد أن :

(ا)احتمال ظهور ٥ أو ٦ = ل (٥ أو ٦)

= ل (٥) + ل (٦) (حدثان متنافيان)

$= \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$

(ب)احتمال ظهور عدد فردى = ل (١ أو ٣ أو ٥)

$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$ (٣ أحداث متنافية)

نتيجة :

إذا كانت $ا_1$ ، $ا_2$ ، ... ، $ا_ن$ هى جميع الأحداث الممكنة فى تجربة ما وكانت هذه الأحداث متنافية فإن مجموع احتمالات هذه الأحداث يساوى الواحد الصحيح ، أى أن :

ل ($ا_1$) + ل ($ا_2$) + ... + ل ($ا_ن$) = ١

لاحظ تحقق هذه النتيجة فى المثال (١ – ٤) السابق .

ثانيا – قاعدة الضرب :

$$ل\,(ا_1\ و ا_2) = ل\,(ا_1)\ .\ ل\,(ا_2 \mid ا_1)\ ،\qquad ل\,(ا_1) \neq 0$$

$$أو\ =\ ل\,(ا_2)\ .\ ل\,(ا_1 \mid ا_2)\ ،\qquad ل\,(ا_2) \neq 0.$$

أى أن احتمال وقوع حدثين معا يساوى احتمال أحدهما مضروبا فى الاحتمال الشرطى للآخر بالنسبة للأول .

حالة خاصة :

إذا كان الحدثان $ا_1$ ، $ا_2$ مستقلين فإنه حسب (٦) تصبح القاعدة كالآتى :

$$ل\,(ا_1\ و ا_2) = ل\,(ا_1)\ .\ ل\,(ا_2)$$

وتسمى هذه النتيجة **بقاعدة الضرب للأحداث المستقلة** ، ويمكن تعميمها كالآتى :

إذا كانت $ا_1$ ، $ا_2$ ، ... ، $ا_ن$ أحداثا مستقلة فإن :

$$ل\,(ا_1\ و ا_2\ و\ ...\ و ا_ن) = ل\,(ا_1)\ .\ ل\,(ا_2)\ ...\ ل\,(ا_ن)$$

مثال (١ – ٩) :

رمى حجرا نرد منتظمان ومتميزان عشوائيا مرة واحدة . نجد أن :

(ا) احتمال ظهور ٥ على القطعة الأولى و٦ على القطعة الثانية

$= ل\,(٥\ و٦) = \frac{1}{٦} \times \frac{1}{٦} = \frac{1}{٣٦}$ **(حدثان مستقلان)**

(ب) احتمال ظهور ٦ على القطعة الأولى و٥ على القطعة الثانية

$= ل\,(٦\ و٥) = \frac{1}{٦} \times \frac{1}{٦} = \frac{1}{٣٦}$ **(حدثان مستقلان)**

(جـ) احتمال ظهور ٥ على إحدى القطعتين و٦ على القطعة الأخرى

$= ل\,(٦\ و٥) + ل\,(٥\ و٦)$ **(حدثان متنافيان)**

$= \frac{1}{٣٦} + \frac{1}{٣٦} = \frac{1}{١٨}$

مثال (١ - ١٠) :

ثلاث مجموعات من الأطفال تتألف الأولى من (٣ بنات ، ولد واحد) وتتألف الثانية من (بنتين ، ولدين) وتتألف الثالثة من (بنت واحدة ، ٣ أولاد) . اختير طفل واحد عشوائياً من كل مجموعة . ما احتمال أن يكون الثلاثة الأطفال المختارون عبارة عن بنت واحدة وولدين ؟

الحـل :

يقع الحدث المطلوب بإحدى الطرق الثلاث الآتية :

(بنت - ولد - ولد) أو (ولد - بنت - ولد) - أو (ولد - ولد - بنت)

احتمال الطريقة الأولى = $\frac{٣}{٤} \times \frac{٢}{٤} \times \frac{٣}{٤} = \frac{٩}{٣٢}$ **(أحداث مستقلة)**

احتمال الطريقة الثانية = $\frac{١}{٤} \times \frac{٢}{٤} \times \frac{٣}{٤} = \frac{٣}{٣٢}$ **(أحداث مستقلة)**

احتمال الطريقة الثالثة = $\frac{١}{٤} \times \frac{٢}{٤} \times \frac{١}{٤} = \frac{١}{٣٢}$ **(أحداث مستقلة)**

وبما أن هذه الطرق الثلاثة متنافية فإن احتمال أن يكون الأطفال المختارون عبارة عن بنت واحدة وولدين هو مجموع احتمالات هذه الطرق ، أي :

$\frac{٩}{٣٢} + \frac{٣}{٣٢} + \frac{١}{٣٢} = \frac{١٣}{٣٢}$

مثال (١ - ١١)

صندوق به ٧ كرات حمراء و٣ كرات بيضاء . سحبت كرتان عشوائياً الواحدة بعد الأخرى دون إرجاع . احسب احتمالات الأحداث الآتية :

(أ) أن تكون كلا الكرتين حمراء . (ب) أن تكون كلا الكرتين بيضاء .

(جـ) أن تكون إحدى الكرتين حمراء والأخرى بيضاء .

الحـل :

(أ) احتمال أن تكون الكرة الأولى حمراء = $\frac{٧}{١٠}$

احتمال أن تكون الكرة الثانية حمراء = $\frac{٦}{٩} = \frac{٢}{٣}$

(هذا احتمال شرطى مع ملاحظة أنه بعد السحبة الأولى يبقى في الصندوق ٩ كرات منها ٦ حمراء) .

احتمال أن تكون كلا الكرتين حمراء = $\frac{٧}{١٠} \times \frac{٢}{٣} = \frac{٧}{١٥}$

وذلك باستخدام قاعدة الضرب وملاحظة أن الاحتمال $\frac{٢}{٣}$ هو احتمال شرطى فهو احتمال ظهور كرة حمراء في السحبة الثانية بشرط أن تكون السحبة الأولى حمراء .

(ب) بنفس المنطق نجد أن :

احتمال أن تكون كلا الكرتين بيضاء = $\frac{٣}{١٠} \times \frac{٢}{٩} = \frac{١}{١٥}$

(جـ) احتمال أن تكون الكرة الأولى حمراء والثانية بيضاء = $\frac{٧}{١٠} \times \frac{٣}{٩} = \frac{٧}{٣٠}$

واحتمال أن تكون الكرة الأولى بيضاء والثانية حمراء = $\frac{٣}{١٠} \times \frac{٧}{٩} = \frac{٧}{٣٠}$

احتمال أن تكون إحدى الكرتين حمراء والأخرى بيضاء = $\frac{٧}{٣٠} + \frac{٧}{٣٠} = \frac{٧}{١٥}$

(حدثان متنافيان)

(٩ – ٨) النماذج الرياضية : MATHEMATICAL MODELS

النموذج لظاهرة ما هو وسيلة لبيان الهيكل العام لتركيب هذه الظاهرة أو للتعبير عن نظرية أو وضع يؤخذ كمولد لما نلمسه من مشاهدات عن هذه الظاهرة . وفي التحليل الإحصائي عادة ما توضع النماذج في صورة رياضية وإن كان هناك صور أخرى بيانية كالخرائط الهندسية والجغرافية والجيولوجية .

والنموذج الرياضي هو تجريد نموذج فيزيائي حيث تحل محل الأشياء والقوى والأحداث رموز تعبر عن متغيرات وبارامترات وثوابت . ويمكن تقسيم النماذج الرياضية إلى ثلاثة أنواع هي : النماذج التحديدية – النماذج الإحصائية – نماذج العمليات العشوائية .

(أ) النماذج التحديدية : DETERMINISTIC MODELS

هى تلك النماذج الرياضية المعتادة التي تستخدم في مختلف المجالات العلمية والتي تعبر عن العلاقة الدالية بين المتغيرات على هيئة قوانين أو معادلات جبرية أو تفاضلية أو تكاملية أو مصفوفية يمكن اشتقاقها نظرياً دون الالتجاء إلى التجريب ، كقوانين نيوتن للحركة .

غير أنه عند التحقق من صحة هذه القوانين تجريبياً نتعرض إلى عدة مصادر للخطأ منها الخطأ في القياس measurement error ومنها خطأ المعادلة equation error أى خطأ استخدام معادلة غير مناسبة ، ومنها الأخطاء العشوائية السابق الإشارة إليها في البند (١ – ٤) . إن هذه الأخطاء لا تدخل في اعتبار النموذج التحديدي ويتطلب تقييمها تحويل هذا النموذج إلى نموذج إحصائي .

(ب) النماذج الإحصائية : STATISTICAL MODELS

النموذج الإحصائي هو تعبير رياضي يشتمل على واحد أو أكثر من المركبات العشوائية بالإضافة إلى المتغيرات والبارامترات والثوابت التي يشتمل عليها النموذج التحديدى . ويمكن اشتقاق نموذج إحصائي من نموذج تحديدى بإضافة صريحة للمركبات العشوائية . فمثلا القانون الشهير الذى يربط المسافة ف والزمن ن لجسيم يسير بسرعة ابتدائية ع. وعجلة منتظمة حـ يأخذ الصورة :

$$ف = ع_. ن + \frac{1}{2} حـ ن^2$$

وهذا نموذج تحديدى . إذا أجرينا تجربة عدة مرات وكانت ف ترمز إلى

المسافة المقطوعة في التجربة الرائية ، نستطيع أن ندخل العامل العشوائي صراحة كالآتي (على أساس أن ع ، ن ، حـ ثوابت) :

$$ف_ر = ع ن + \frac{1}{2} حـ ن^{2} + خ_ر$$

$$أو ف_ر = ف + خ_ر$$

حيث ف هى القيمة الحقيقية للمسافة كما تحسب من النموذج التحديدى ، $خ_ر$ هى انحراف المسافة الناتجة في التجربة الرائية عن المسافة الحقيقية ف .

وسنلقى أمثلة كثيرة للنماذج الإحصائية خاصة عند دراسة تحليل التباين وتحليل الانحدار .

على أنه ليس من الضرورى أن يعبر النموذج الإحصائي عن نموذج فيزيائي بالوضوح الظاهر في القانون السابق بل يمكن اختيار نموذج إحصائي مناسب اعتماداً على ما نتصوره من مصادر تؤثر فيما نشاهده من الاختلافات في البيانات الناتجة عن الظواهر التي ندرسها .

(جـ) نماذج العمليات العشوائية : STOCHASTIC PROCESS MODELS

هذا النوع من النماذج قد يشتمل على عوامل عشوائية مماثلة لتلك التي في النماذج الإحصائية ولكنه بالاضافة إلى ذلك يشتمل على عملية عشوائية معينة تدخل في بناء النموذج وتصف الظاهرة على أساس احتمالى وليس على أساس تحديدى . ومن هذه النماذج النموذج IIالذى سنقدمه فى فصل تحليل التباين بالبند (٨ — ١٤) .

تمـــارين (١)

١ – الآتى بيان الأعمار بالسنوات للأطفال الذين ذهبوا إلى إحدى المستشفيات للعلاج في أحد الأيام وعددهم ٢٤٠ .

(أ) أعط هذه الأعمار أرقاماً مسلسلة من ٠٠١ إلى ٢٤٠

(ب) استخدم جدول الأعداد العشوائية لاستخراج عينة عشوائية حجمها ٣٠ (اذكر رقم العمود ورقم الصف اللذين بدأت بهما) .

(جـ) أوجد الوسط الحسابي لأعمار أفراد هذه العينة .

(د) كرر الخطوتين (ب) ، (جـ) عدة مرات وقارن بين الأوساط الحسابية للعينات الناتجة وبين الوسط الحسابي لأعمار الأطفال جميعاً .

٧	١٠	٣	١٠	٨	٢
٥	١٤	٧	١٣	١٢	٥
٧	١١	٦	١١	٢	٥
٦	٢	٢	٥	١	٥
١٣	٥	٤	٣	٢	٩
٢	٣	١	١	٤	٤
٢	١	٦	١٥	١	١٣
٥	٥	٦	٢	٥	٨
٨	٢	٩	١٤	١٠	٥
٤	٣	١٤	٦	٧	١٤
٣	٨	١	٢	٨	٥
٤	١	٢	٤	١	١
٩	٢	١	١	١٥	١٣
٣	٥	٤	٣	٥	٨
٢	٦	١١	٧	٣	١٤

١١	٩	١٠	٨	٣	١
٨	١	١	١	٦	٣
٢	٥	٥	١	٨	١٥
١	١٤	٩	٢	١	٦
٤	١٠	٦	٥	١٢	٢
١٥	٤	١	١١	٢	٢
٧	٩	٧	٥	١٠	١١
١	١٥	٢	٢	٢	٦
١٣	٤	١	٢	٥	٤
١٠	١	٣	١٣	١٥	١٤
٦	٣	٤	٩	٢	٣
٤	٨	٧	٥	٢	١
١	١١	٣	١٣	١٢	٦
١٤	١	١١	٣	٩	٥
١٠	٤	٢	٥	١	١٠
٣	٧	٣	٩	١٤	٢
١٢	١٣	٩	١	١٠	٩
٥	١	٥	١٤	٥	٢
١٢	٧	٦	٣	١٢	١٣
٣	٣	١	٣	٢	٣
٢	٩	١٢	٨	١٠	٧
٩	٢	٩	٥	٥	٤
٤	٤	٣	٢	١	٧
١	١	١	١٢	٥	١٢
٨	١٠	٧	٧	١٤	٣

٢ – قرّب الأعداد الآتية إلى درجة الدقة المشار إليها .

٦٥,٤ إلى أقرب وحدة ، ٠,٠٦٣٥ إلى أقرب جزء من ألف

١٤٤,٥ إلى أقرب وحدة ، ٢,٥٠٠١ إلى أقرب وحدة

٣,٥٤٢ إلى أقرب جزء من مائة ، ٢٤,٦٢ إلى رقمين معنويين .

٣ – اجمع الأعداد ٨,٣٤ ، ٦,١٢ ، ٧,٥٤ ، ٦,٥٦ ، ٦,٤٧ ، ٧,٢٨ ، ٧,١٧

(أ) مباشرة (ب) بعد تقريب كل منها إلى جزء من عشرة .

٤ – صندوق به ٦ كرات حمراء و٤ كرات صفراء . سحبنا منه عشوائياً كرتين الواحدة بعد الأخرى دون إعادة .

(أ) احسب احتمال أن تكون كلا الكرتين حمراء .

(ب) احسب احتمال أن تكون كلا الكرتين صفراء .

(جـ) احسب احتمال أن تكون واحدة حمراء وواحدة صفراء .

٥ – مجموعتان من الأطفال تتألف الأولى من (٣ بنات ، ولدين) وتتألف الثانية من (بنتين ، ٣ أولاد) . اختير طفل واحد عشوائيا من كل مجموعة . ما احتمال أن يكون الطفلان المختاران عبارة عن بنت واحدة وولد واحد ؟ .

الفصل الثانى

التوزيعات التكرارية

FREQUENCY DISTRIBUTIONS

إن البيانات التي نحصل عليها من عينة ما عن متغير ما تكون عبارة عن عدد من القيم أو القراءات مسجلة كيفما اتفق ، ولذلك تسمى عادة بالبيانات الخام raw data . وأول ما نفعله بهذه البيانات هو تنظيمها وتلخيصها في صورة مركزة عادة ما تكون على هيئة توزيعات تكرارية موضوعة فى جداول مناسبة ، وكثيراً ما نقوم بتمثيلها بيانياً . إن هذا التنظيم يجعل البيانات أكثر طواعية للدراسة والتحليل وقد يكشف عن بعض الصفات البارزة أو الخصائص العامة التي لا تظهر في القراءات قبل تنظيمها . ولا مفر لأى دارس للإحصاء من أن يكون على دراية بأمور أساسية ثلاثة هي :

(أ) كيفية إنشاء الجداول التكرارية .
(ب) كيفية تمثيل التوزيعات بيانيا .
(جـ) كيفية وصف التوزيعات وصفاً موضوعياً .

وهذا ما نتناوله هنا بالتلخيص المركز عن طريق الأمثلة ، على أساس أن القارىء سبق له دراستها .

(٢ – ١) الجداول التكرارية : FREQUENCY TABLES

(٢ – ١ – ١) جدول التوزيع التكرارى البسيط :

مثال (٢ – ١) :

الأعداد الآتية تعبر عن النسب المئوية للكاربون الذى وجد في عينة عشوائية حجمها ٢٥ في نوع من الفحم .

٨٠ ٨١ ٨٦ ٨٧ ٨٥ ٨٦ ٨٧ ٧٧ ٨٢ ٨٠ ٨٣ ٨٤ ٨٦
٨٣ ٨٥ ٨٤ ٨٦ ٨٥ ٧٨ ٨٢ ٧٩ ٧٩ ٨٦ ٧٧ ٨٧

كون الجدول التكرارى البسيط وجدول التكرارات المتجمعة المئوية .

الحـل :

ننشيء جدولا كالجدول (٢ – ١) التالى حيث يشتمل العمود الأول منه على القيم المختلفة للمتغير مرتبة ترتيباً تصاعدياً (أو تنازلياً) ويشتمل العمود الثاني منه على عدد من الشرط أمام كل قيمة بالعمود الأول تحصي عدد مرات وجود هذه القيمة بالبيانات الخام . أما العمودان الثالث والرابع فأمرهما واضح وكذلك بالنسبة للجدول (٢ – ٢) .

الجدول (٢ – ١)
التكرارات والتكرارات النسبية
للنسب المئوية للكاربون في عينة من الفحم

س ر	الشرط	ك ر	ح ر
٧٧	//	٢	٠,٠٨
٧٨	/	١	٠,٠٤
٧٩	//	٢	٠,٠٨
٨٠	//	٢	٠,٠٨
٨١	/	١	٠,٠٤
٨٢	//	٢	٠,٠٨
٨٣	//	٢	٠,٠٨
٨٤	//	٢	٠,٠٨
٨٥	///	٣	٠,١٢
٨٦	//// (مشطوبة)	٥	٠,٢٠
٨٧	///	٣	٠,١٢
المجموع		٢٥	١,٠٠

الجدول (٢ – ٢)
التكرارات المتجمعة والمتجمعة المئوية

الحدود العليا	التكرار المتجمع	التكرار المتجمع ٪
> ٧٧	٢	٨
> ٧٨	٣	١٢
> ٧٩	٥	٢٠
> ٨٠	٧	٢٨
> ٨١	٨	٣٢
> ٨٢	١٠	٤٠
> ٨٣	١٢	٤٨
> ٨٤	١٤	٥٦
> ٨٥	١٧	٦٨
> ٨٦	٢٢	٨٨
> ٨٧	٢٥	١٠٠

ملاحظات :

(١) $س_ر$ ترمز إلى القيم المختلفة للمتغير .

(٢) $ك_ر$ ترمز إلى تكرار القيمة $س_ر$ أى إلى عدد مرات وجود هذه القيمة بالبيانات الخام .

(٣) $ح_ر$ ترمز إلى التكرار النسبي للقيمة $س_ر$ أى خارج قسمة العدد $ك_ر$ على حجم التوزيع وهو هنا ٢٥ . وهذه التكرارات النسبية تؤخذ كتقديرات للاحتمالات تحت شروط معينة منها عشوائية العينة وكبر حجمها .

(٤) تسمى مجموعة الأزواج المرتبة

{ (٧٧ ، ٢) ، (٧٨ ، ١) ، ... ، (٨٧ ، ٣) }

بالتوزيع التكرارى للمتغير س في العينة ، وهذه المجموعة تشكل العمودين الأول والثالث من الجدول (٢ – ١) .

(٢ – ١ – ٢) جدول التوزيع التكرارى المجمع في فئات :

حين تشتمل البيانات على عدد كبير من قيم متغير عددى ، يفضل تجميع هذه القيم في فئات فتوضع كل مجموعة من القيم المتقاربة في فئة خاصة ، ويراعى هنا ألا يكون عدد هذه الفئات كبيراً فتنتفي الحكمة أو الفائدة من عملية التجميع ، وألا يكون عددها صغيراً فتضيع معالم التوزيع ويفقد الكثير من تفاصيله .

مثال (٢ – ٢) :

قيست أطوال محيطات الرؤوس بالمليمترات لعينة حجمها ٤٠ من الحمام المنزلى فوجدت كما يلى :

١٢,٢ ١٢,٩ ١١,٨ ١١,٩ ١١,١ ١٢,٣ ١٢,٢ ١١,٨ ١١,٨ ١١,٦

١٠,٧ ١١,٥ ١١,٣ ١١,٢ ١١,٦ ١١,٩ ١٣,٣ ١١,٢ ١٠,٥ ١١,١

١٢,١ ١١,٩ ١٠,٤ ١٠,٧ ١٠,٨ ١١,٠ ١١,٩ ١٠,٢ ١٠,٩ ١١,٦

١٠,٨ ١١,٦ ١٠,٤ ١٠,٧ ١٢,٠ ١٢,٤ ١١,٧ ١١,٨ ١١,٣ ١١,١

جمع هذه البيانات في توزيع تكراري ذي فئات وأوجد توزيع التكرارات المتجمعة النسبية المئوية .

الحـل :

المدى = أكبر قيمة للمتغير – أصغر قيمة للمتغير

= ١٣,٣ – ١٠,٢ = ٣,١

إذا رأينا أن نأخذ حوالى ١٠ فئات يكون طول كل فئة ٣,١ ÷ ١٠ = ٠,٣٠ تقريباً .

وعلى أساس أن الأطوال قيست لأقرب جزء من عشرة من المليمتر سنعتبر أنه إذا كان س هو العدد الذى سجلناه لطول محيط رأس حمامة فإن الطول الحقيقي لهذا الرأس يقع بين العددين س ± ٠,٠٥ . فمثلا أصغر عدد مسجل هو ١٠,٢ وإذن الطول الحقيقي لمحيط رأس أصغر حمامة في العينة يقع بين العددين ١٠,٢ ± ٠,٠٥ أى بين العددين ١٠,١٥ ، ١٠,٢٥ .

نأخذ العدد ١٠,١٥ كحد أدنى للفئة الأولى . وحيث أننا اخترنا أن يكون طول الفئة ٠,٣٠ ، فإن الحد الأعلى لهذه الفئة يكون ١٠,١٥ + ٠,٣٠ = ١٠,٤٥ ويكون هذا العدد نفسه هو الحد الأدنى للفئة الثانية التي ينبغى أن يكون حدها الأعلى ١٠,٤٥ + ٠,٣٠ = ١٠,٧٥ ونستمر في إنشاء الفئات التالية بنفس الطريقة حتى نصل إلى فئة تغطى أكبر عدد في البيانات المعطاة وهو ١٣,٣ — انظر العمود الأول من الجدول (٢ – ٣) الآتي . ويلاحظ أن هذا الأسلوب في تكوين الفئات للقياسات المقربة يسمح بأن يكون لكل عدد في البيانات المعطاة مكان وحيد في إحدى هذه الفئات . وبالنسبة لملأ العمود الثالث نمر على الأعداد المعطاة واحداً واحداً ونضع شرطة أمام الفئة التي يدخل فيها .

الجدول (٢ - ٣)

التكرارات والتكرارات النسبية لمحيطات الرؤوس

بالمليمترات لعينة الحمام المنزلى

الفئات	س ر	الشرط	ك ر	ح ر
١٠,٤٥-١٠,١٥	١٠,٣	///	٣	٠,٠٧٥
١٠,٧٥-١٠,٤٥	١٠,٦	////	٤	٠,١١٠
١١,٠٥-١٠,٧٥	١٠,٩	////	٤	٠,١٠٠
١١,٣٥-١١,٠٥	١١,٢	// ~~////~~	٧	٠,١٧٥
١١,٦٥-١١,٣٥	١١,٥	~~////~~	٥	٠,١٢٥
١١,٩٥-١١,٦٥	١١,٨	//// ~~////~~	٩	٠,٢٢٥
١٢,٢٥-١١,٩٥	١٢,١	////	٤	٠,١٠٠
١٢,٥٥-١٢,٢٥	١٢,٤	//	٢	٠,٠٥٠
١٢,٨٥-١٢,٥٥	١٢,٧	—	٠	٠,٠٠٠
١٣,١٥-١٢,٨٥	١٣,٠	/	١	٠,٠٢٥
١٣,٤٥-١٣,١٥	١٣,٣	/	١	٠,٠٢٥
المجموع			٤٠	١,٠٠

الجدول (٢ - ٤)

التكرارات المتجمعة

والمتجمعة المئوية

الحدود العليا	التكرار المتجمع	التكرار المتجمع ٪
≥ ١٠,٤٥	٣	٧,٥
≥ ١٠,٧٥	٧	١٧,٥
≥ ١١,٠٥	١١	٢٧,٥
≥ ١١,٣٥	١٨	٤٥,٥
≥ ١١,٦٥	٢٣	٥٧,٥
≥ ١١,٩٥	٣٢	٨٠,٠
≥ ١٢,٢٥	٣٦	٩٠,٠
≥ ١٢,٥٥	٣٨	٩٥,٠
≥ ١٢,٨٥	٣٨	٩٥,٠
≥ ١٣,١٥	٣٩	٩٧,٥
≥ ١٣,٤٥	٤٠	١٠٠,٠

ملاحظات :

١ - س ر ترمز إلى مركز الفئة **class mark** وهو الوسط الحسابي لحدى الفئة ، فمثلا مركز الفئة الأولى هو $\frac{1}{2}$ (١٠,١٥ + ١٠,٤٥) = ١٠,٣ .

ويؤخذ مركز الفئة ممثلا لها بمعني أننا نعتبر أن جميع القيم التي دخلت الفئة مساوية لهذا المركز ، فمثلا تضم الفئة الأولى (١٠,١٥ - ١٠,٤٥) ثلاثة من الأعداد المعطاة هى ١٠,٤ ، ١٠,٢ ، ١٠,٤ غير أننا في عملية التجميع نلغى هذه الأعداد ونعتبر أن بهذه الفئة ثلاثة أعداد كل منها يساوى مركز الفئة وهو ١٠,٣ .

كذلك تضم الفئة الثانية أربعة أعداد هى ١٠,٧ ، ١٠,٥ ، ١٠,٧ ، ١٠,٧ غير أننا نعتبر أن بهذه الفئة أربعة أعداد كل منها يساوى مركز الفئة وهو ١٠,٦ . وفي اعتبارنا هذا شيء من التجاوز يسمى بخطأ التجميع ، إلا أن هذه الأخطاء عادة ما يلغى بعضها البعض لأن بعضها بالزيادة والبعض الآخر بالنقصان ، ولاسيما إذا كان حجم التوزيع كبيراً .

٢ – في تكوين الفئات في هذا المثال راعينا أن المتغير هو متغير عددى من النوع المتصل وأن القياس كان إلى أقرب جزء من عشرة من المليمتر . أما إذا اعتبرنا أن القياس مضبوط فيمكن أن نضع الفئات كالآتي :

١٠,٢ – لتعني الفئة التي تشمل الأعداد بدءاً من ١٠,٢ إلى أقل من ١٠,٥
١٠,٥ – لتعني الفئة التي تشمل الأعداد بدءاً من ١٠,٥ إلى أقل من ١٠,٨
١٠,٨ – لتعني الفئة التي تشمل الأعداد بدءاً من ١٠,٨ إلى أقل من ١١,١
وهكذا .. .. ..

وتستخدم هذه الطريقة أيضاً حين يكون المتغير من النوع الوثاب . ولبيان أن هذه الطريقة لا تصلح فى الحالة التى تكون فيها البيانات مسجلة بمقياس تقريبى ، اعتبر الحمامة التى سجل طولها على أنه ١٠,٥ ملليمترا (تقريبا) . نعلم أن الطول الحقيقى لهذه الحمامة يقع بين العددين ١٠,٤٥ ، ١٠,٥٥ وعلى ذلك فإن الطول الحقيقى قد يكون أصغر من الطول المسجل ١٠,٥ ، مثلا ١٠,٤٨ ، وفى هذه الحالة ينبغى وضعه فى الفئة ١٠,٢ – أو قد يكون أكبر من ١٠,٥ ، مثلا ١٠,٥٤ ، وفى هذه الحالة ينبغى وضعه فى الفئة ١٠,٥ — وما دمنا لا نعرف الطول الحقيقى لهذه الحمامة فإننا نكون فى حيرة من استخدام أى من هاتين الفئتين . ونقع فى هذه الحيرة أيضا فى تناول كثير من الأطوال الأخرى مثل ١٠,٨ ، ١١,١ ، ١١,٤ ، ... ومن هذا نرى أن هذه الطريقة لا تضمن أن يكون لكل قيمة (من القيم المقربة) مكان فى واحدة وواحدة فقط من الفئات .

(٢ – ١ – ٣) الجدول التكراري المزدوج (أو جدول الاقتران) :

كل من المثالين السابقين يتناول توزيعاً تكرارياً لمتغير واحد ، وفيما يلي مثالان يتناول كل منهما التوزيع التكراري المشترك لمتغيرين . joint distribution

مثال (٢ – ٣) :

الجدول (٢ – ٥) الآتي يعطي التكرارات المشاهدة لطول محيط الرأس وطول الطفل ساعة الولادة في عينة من ٩٩ مولوداً .

الجدول (٢ – ٥)

محيط الرأس	طول الجسم			المجموع
	٤٧ –	٥٠ –	٥٣ –	
٣٢ –	٤٠	٣٩	٢	٧٨
٣٦ –	صفر	١٤	٧	٢١
المجموع	٤٠	٥٠	٩	٩٩

لدينا متغيران هما (١) طول محيط الرأس وقد قسمت الأطوال إلى فئتين (٢) طول الجسم وقد قسمت الأطوال إلى ثلاث فئات ، ولهذا يسمى مثل هذا الجدول بجدول اقتران ٢ × ٣ 2 x 3 contingency table لأن المتغيرين يقترنان فيه في توزيع مشترك .

من هذا الجدول نستطيع استخراج الجدولين (٢ – ٦) ، (٢ – ٧) الآتيين :

الجدول (٢ – ٦)
التوزيع الهامشى لطول محيط الرأس

محيط الرأس	ك ر
٣٢ –	٧٨
٣٦ –	٢١
المجموع	٩٩

الجدول (٢ – ٧)
التوزيع الهامشي لطول الطفل

طول الجسم	ك ر
٤٧ –	٤٠
٥٠ –	٥٠
٥٣ –	٩
المجموع	٩٩

يعطى الجدول (٢ – ٦) ما يسمى بالتوزيع الهامشى للمتغير الأول (طول محيط الرأس) وهو يعني التوزيع التكرارى لهذا المتغير بصرف النظر عن المتغير الثاني . وبالمثل يعطى الجدول (٢ – ٧) التوزيع الهامشي للمتغير الثاني (طول الجسم) .

مثال (٢ – ٤) :

في إحدى التجارب قسم ١٤٦٩ من الرجال في الأعمار ما بين ٦٠ ، ٦٤ عاماً من حيث عادة التدخين إلى قسمين : يدخن ولا يدخن . وبعد ٦ سنوات من بدء التجربة حسب عدد الوفيات للقسمين فنتج التوزيع التكرارى المزدوج المبين بالجدول (٢ – ٨) وهو يعطى التوزيع التكرارى المشترك لمتغيرين من النوع الوصفي هما الوفاة وعادة التدخين .

مثل هذا الجدول يسمى بجدول اقتران ٢ × ٢ لأن كلا من المتغيرين مقسم إلى قسمين . استخرج التوزيع الهامشي لكل من المتغيرين .

الجدول (٢ – ٨)

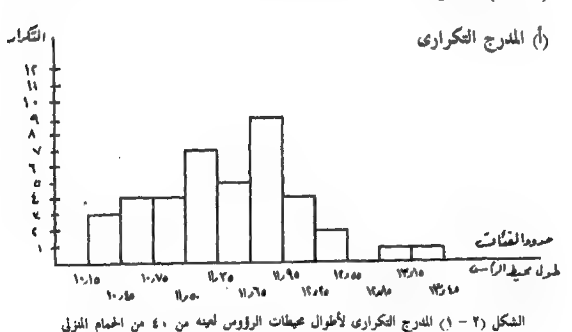
التوزيع المشترك لخاصتي الوفاة والتدخين لعينة من كبار السن

الوفاة	التدخين: يدخن	التدخين: لا يدخن	المجموع
توفي	٥٤	١١٧	١٧١
حي	٣٤٨	٩٥٠	١٢٩٨
المجموع	٤٠٢	١٠٦٧	١٤٦٩

(٢ – ٢) التمثيل البياني للتوزيعات التكرارية :

المعتاد في تمثيل التوزيعات بيانيا إنشاء محورين متعامدين في المستوى يجزأ كل منهما بمقياس رسم مناسب بحسب الصورة البيانية التي نرغب في تقديمها ، والأشكال الثلاثة الآتية تعرض أشهر هذه الصور وهى تمثل البيانات الواردة بالمثال (٢ – ٢) السابق .

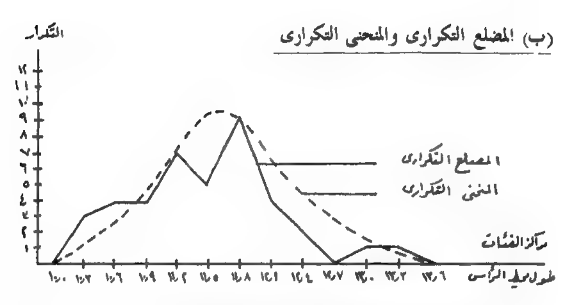
(أ) المدرج التكرارى



الشكل (٢ – ١) المدرج التكرارى لأطوال محيطات الرؤوس لعينه من ٤٠ من الحمام المنزلى

هذا الشكل يعطى ما يسمى بالمدرج التكرارى (هستوجرام) histogram وهو يؤخذ من العمودين الأول والرابع من الجدول (٢ – ٣) ويتألف من عدد من المستطيلات المتلاصقة قواعدها فئات التوزيع وارتفاعاتها تتناسب مع التكرارات المناظرة .

(ب) المضلع التكرارى والمنحنى التكرارى



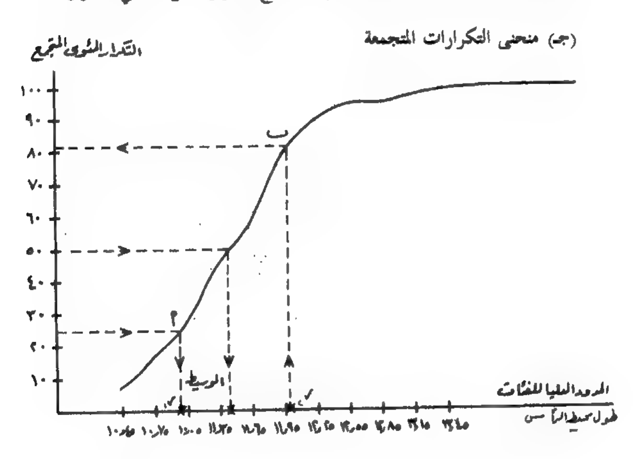
الشكل (٢ – ٢) المضلع التكرارى والمنحني التكرارى لأطوال محيطات الرؤوس لعينة من ٤٠ من الحمام المنزلى

هذا الشكل يعطى ما يسمى بالمضلع التكرارى frequency polygon وهو يؤخذ من العمودين الثاني والرابع من الجدول (٢ – ٣) . يمثل المحور الأفقي مراكز الفئات ويمثل المحور الرأسي التكرارات وينتج المضلع من توصيل عدد من النقط (١١ نقطة) إحداثياتها الأفقية مراكز الفئات وإحداثياتها الرأسية التكرارات المناظرة ثم يغلق المضلع من الطرفين وذلك بتصور وجود فئة إضافية في بداية التوزيع وفئة إضافية في آخره التكرار في كل منهما هو بطبيعة الحال صفر .

كما يعطى هذا الشكل ما يسمى بالمنحني التكرارى frequency curve وهو منحني ناعم يمهد باليد ماراً ببعض هذه النقط وقريباً من البعض الآخر ، أى ليس

من الضروری أن یمر بها جمیعاً لأن الهدف من رسمه هو محاولة استکشاف الاتجاه العام لتوزيع المتغير في المجتمع الذى أخذت منه العينة ومن الواضح أن عملية التمهيد هذه تعتمد على ذاتية الراسم وتختلف من شخص إلى آخر ، وهى تجرى على أساس أن التوزيع التكرارى الذى لدينا هو توزيع لعينة مأخوذة من مجتمع متصل ، وكلما كبر حجم العينة وصغرت أطوال الفئات كلما اقترب المضلع التكرارى من المنحني التكرارى .

(جـ) منحنى التكرارات المتجمعة



الشكل (٢ – ٣) منحنى التكرارات المتجمعة المئوية لأطوال محيطات الرؤوس لعينة من الحمام المنزل .

هذا الشكل يعطى منحني التكرارات المتجمعة المئوية percentage cumulative frequency curve . وهو يؤخذ من العمودين الأول والثالث من الجدول (٢ – ٤) أى أن الإحداثيات الأفقية للنقط هى الحدود العليا للفئات والإحداثيات الرأسية هى التكرارات المتجمعة المئوية المناظرة . وكان من الممكن أن ترسم المنحني نفسه من

العمودين الأول والثاني إلا أن هذا يحتاج إلى التفكير في مقياس رسم مناسب لكل توزيع على حدة ، أما استخدام التكرارات المتجمعة المئوية فمن شأنه أن يكون تقسيم المحور الرأسي ثابت لأى توزيع .

ومن هذا المنحنى نستطيع الإجابة إجابة تقريبية عن نوعين من الأسئلة يتمثلان فيما يلى :

(أ) ما طول محيط رأس الحمامة الذى يقل عنه أو يساويه ٢٥٪ من أطوال محيطات رؤوس الحمام ؟

(ب) ما النسبة المئوية لعدد الحمام الذى تقل أو تساوى أطوال محيطات رؤوسها عن ١٢ ملليمترا ؟

وللإجابة عن السؤال الأول نرسم من النقطة التي تمثل التكرار المتجمع المئوى ٢٥٪ على المحور الرأسي خطا مستقيما يوازى المحور الأفقي ويقطع المنحني في نقطة (أ) ثم نرسم من (أ) خطا مستقيما يوازى المحور الرأسي يلقي المحور الأفقي عند النقطة $م_١$. وبعملية حسابية بسيطة نجد أن $م_١$= ١٠,٩٨ تقريبا فيكون الطول المطلوب هو ١٠,٩٨ ملليمترا تقريبا . أما الإجابة عن السؤال الثاني فتسير بعكس خطوات الإجابة عن السؤال الأول فنرسم من النقطة التي تمثل العدد ١٢ على المحور الأفقي مستقيما يوازى المحور الرأسي ويقطع المنحني في نقطة ب ثم نرسم من ب مستقيما يوازى المحور الأفقي ليلقي المحور الرأسي عند النقطة ٨٢ تقريبا فتكون النسبة المطلوبة هى ٨٢٪ تقريبا .

ومن منحني التكرارات المتجمعة المئوية نستطيع بنفس الطريقة أن نوجد تقريبيا ما يسمى بالمئينات والربيعات وهى أعداد تستخدم في وصف التوزيعات كما سنرى بعد . وهى من المقاييس المسماة بمقاييس الموضع تمييزها عن مقاييس المقدار التى سنقدمها فى البند (٢ — ٤) .

(٢ – ٣) المئينات والربيعات CENTILES AND QUARTILES

إذا كان لدينا توزيع تكرارى لمتغير كمى رتبت قيمه تصاعديا فإن المئينات $م_١$ ، $م_٢$ ، .... ، $م_{٩٩}$ لهذا التوزيع تعرف بأنها تلك الأعداد التي تقسمه إلى ١٠٠ قسم يشتمل كل منها على عدد متساوى من قيم المتغير أى على ١٪ من هذه القيم . وعلى ذلك فالمئين $م_{١٠}$ هو العدد الذى يقل عنه أو يساويه ١٠٪ من قيم المتغير والمئين $م_{٨٠}$ هو العدد الذى يقل عنه أو يساويه ٨٠٪ من قيم المتغير وهكذا .

وبالمثل الربيعات $ر_١$ ، $ر_٢$ ، $ر_٣$ للتوزيع تعرف بأنها تلك الأعداد التي تقسمه إلى أربعة أقسام يشتمل كل منها على ربع قيم المتغير . ويلاحظ أن :

الربيع الأول $ر_١$ = المئين $م_{٢٥}$ = العدد الذى يقل عنه أو يساويه ٢٥٪ من قيم المتغير

= ١٠,٩٨ مليمترا تقريبا في هذا المثال .

الربيع الثاني $ر_٢$ = المئين $م_{٥٠}$ = العدد الذى يقل عنه أو يساويه ٥٠٪ من قيم المتغير .

= ١١,٥ مليمترا تقريبا في هذا المثال .

ويسمى هذا المقياس أيضا **بالوسيط** لأنه يتوسط التوزيع ويقسمه إلى قسمين متساويين في عدد قيم المتغير .

الربيع الثالث $ر_٣$ = المئين $م_{٧٥}$ = العدد الذى يقل عنه أو يساويه ٧٥٪ من قيم المتغير .

= ١١,٨٨ ملليمترا تقريبا في هذا المثال .

وفي المثال (٢ – ٢) حيث ن = ٤٠ حمامة نجد أن الربيع الأول وهو ١٠,٩٨ يسبقه عشرة أعداد تقع قيمها من ١٠,٢ إلى ١٠,٩ ، كما نجد أن الوسيط وهو

١١,٥ يسبقة عشرون عددا تقع قيمها من ١٠,٢ إلى ١١,٥ ، ونجد أن الربيع الثالث وهو ١١,٨٨ يسبقه ثلاثون عددا تقع قيمها من ١٠,٢ إلى ١١,٨ .

ونستطيع إيجاد المئينات والربيعات بطريقة حسابية وهى طريقة أكثر دقة نوضحها عن طريق التوزيع الذى بالمثال (٢ - ٢) والذى يمكن تلخيصه بالجدول (٢ - ٩) الآتي :

الجدول(٢ - ٩)

التوزيع التكرارى والتوزيع المتجمع لأطوال محيطات الرؤوس بالمليمترات لعينة من الحمام المنزلى

الفئـــة	التكرار	التكرار المتجمع
١٠,١٥ — ١٠,٤٥	٣	٣
١٠,٤٥ — ١٠,٧٥	٤	٧
١٠,٧٥ — ١١,٠٥	٤	١١
١١,٠٥ — ١١,٣٥	٧	١٨
١١,٣٥ — ١١,٦٥	٥	٢٣
١١,٦٥ — ١١,٩٥	٩	٣٢
١١,٩٥ — ١٢,٢٥	٤	٣٦
١٢,٢٥ — ١٢,٥٥	٢	٣٨
١٢,٥٥ — ١٢,٨٥	٠	٣٨
١٢,٨٥ — ١٣,١٥	١	٣٩
١٣,١٥ — ١٣,٤٥	١	٤٠

لايجاد الوسيط $س_٢$ والربيعين $س_١$ ، $س_٣$:

ترتيب الوسيط = $\frac{ن}{٢}$ (هذه قاعدة عامة لتوزيعات المتغيرات المتصلة).

= $\frac{٤٠}{٢}$ = ٢٠

إذن الوسيط هو الطول الذى يقل عنه أو يساويه أطوال ٢٠ حمامة .

نلاحظ من الجدول أن هناك ١٨ حمامة تقل أطوالها عن ١١,٣٥ مليمترا .

وأن هناك ٢٣ حمامة تقل أطوالها عن ١١,٦٥ مليمترا .

وعلى ذلك يجب أن يقع الوسيط في الفئة ١١,٣٥ – ١١,٦٥ وتسمى هذه الفئة حينئذ **بالفئة الوسيطية** . إن هذه الفئة تشتمل على ٥ وحدات (حمامات)نريد أن ناخذ منها اثنين فقط لنستكمل العدد المطلوب (من ١٨ حمامة إلى ٢٠ حمامة) وعلى فرض أن الوحدات الخمسة موزعة بانتظام داخل الفئة بمعني أنها تبعد عن بعضها بمسافات متساوية فإن العدد ٢ يمثل طولا قدره $\frac{٢}{٥}$ طول الفئة أى طولا قدره $\frac{٢}{٥}$ × ٠,٣ = ٠,١٢ م ولما كانت بداية الفئة ١١,٣٥ م فإن :

الوسيط = $س_٢$ = ١١,٣٥ + ٠,١٢ = ١١,٤٧ مليمترا .

وبصفة عامة نوجد الوسيط من الصيغة الآتية :

الوسيط = الحد الأدنى للفئة الوسيطية +

$$\frac{\text{ترتيب الوسيط – التكرار المتجمع للفئة السابقة للفئة الوسيطية}}{\text{التكرار في الفئة الوسيطية}} \times \text{طول الفئة}$$

وبنفس المنطق السابق نوجد الربيعين الأول والثالث بالصيغتين الآتيتين مع ملاحظة أن ترتيب الربيع الأول هو $\frac{ن}{٤}$ وترتيب الربيع الثالث هو $\frac{٣ن}{٤}$ وذلك بالنسبة للتوزيعات التكرارية للمتغيرات المتصلة .

$ر_1$ = الحد الأدنى لفئة الربيع الأول +

$$\frac{\text{ترتيب } ر_1 - \text{التكرار المتجمع للفئة السابقة لفئة الربيع الأول}}{\text{التكرار في فئة الربيع الأول}} \times \text{طول الفئة}$$

$ر_3$ = الحد الأدنى لفئة الربيع الثالث +

$$\frac{\text{ترتيب } ر_3 - \text{التكرار المتجمع للفئة السابقة لفئة الربيع الثالث}}{\text{التكرار في فئة الربيع الثالث}} \times \text{طول الفئة}$$

فى المثال (٢ – ٢) نجد ما يلى :

ترتيب الربيع الأول = $\frac{٤٠}{٤}$ = ١٠

$ر_1$ = ١٠,٧٥ + $\frac{١٠-٧}{٤}$ × ٠,٣ = ١٠,٧٥ + ٠,٢٢٥

= ١٠,٩٧٥ مليمترا .

، ترتيب الربيع الثالث = $\frac{٤٠ \times ٣}{٤}$ = ٣٠

$ر_3$ = ١١,٦٥ + $\frac{٣٠-٢٣}{٩}$ × ٠,٣ = ١١,٦٥ + ٠,٢٣٣

= ١١,٨٨٣ مليمترا .

وتطبق هذه الصيغ أيضا في حالة التوزيعات التكرارية البسيطة وفي حالة التوزيعات غير التكرارية أى التي على صورة مجموعة من القيم $س_1$ ، $س_2$ ، ... ، $س_ن$ (تكرار كل منها الواحد الصحيح) ، بشرط أن يكون المتغير من النوع المتصل فتعتبر أى قيمة س من قيم المتغير كأنها فترة تبدأ بنصف وحدة خطوة أسفل العدد س وتنتهى بنصف وحدة خطوة أعلاه وبذلك يكون طول الفترة مساويا الواحد إذا كانت س مقدرة بالوحدات الصحيحة ويساوى ٠,١ إذا كانت مقدرة إلى خانة عشرية واحدة ويساوى ٠,٠١ إذا كانت مقدرة إلى خانتين عشريتين وهكذا . ونوضح ذلك بأخذ المثال (٢ – ١) الذى نلخصه بالجدول (٢ – ١٠) الآتى :

الجدول (٢ – ١٠)

س	ك	≥ س
٧٧	٢	٢
٧٨	١	٣
٧٩	٢	٥
٨٠	٢	٧
٨١	١	٨
٨٢	٢	١٠
٨٣	٢	١٢
٨٤	٢	١٤
٨٥	٣	١٧
٨٦	٥	٢٢
٨٧	٣	٢٥

ترتيب الربيع الأول = ٢٥/٤ = ٦,٢٥

الربيع الأول يقع في الفئة التي تعبر عن العدد ٨٠ التي تمتد من ٧٩,٥ إلى ٨٠,٥ والتكرار فيها ٢

ر١ = ٧٩,٥ + (٦,٢٥ − ٥)/٢ × ١ = ٨٠,١٢٥

، ترتيب الوسيط = ٢٥/٢ = ١٢,٥

الوسيط يقع في الفئة التي تعبر عن العدد ٨٤ التي تمتد من ٨٣,٥ إلى ٨٤,٥ والتكرار فيها ٢

ر٢ = ٨٣,٥ + (١٢,٥ − ١٢)/٢ × ١ = ٨٣,٧٥

، ترتيب الربيع الثالث = (٣×٢٥)/٤ = ١٨,٧٥

ر٣ = ٨٥,٥ + (١٨,٧٥ − ١٧)/٥ × ١ = ٨٥,٨٥

ملاحظـــة : لا تصلح هذه الطريقة ولا طريقة المنحني المتجمع في حالة توزيعات المتغيرات غير المتصلة (الوثابة) ، على أنه في حالة التوزيعات غير التكرارية من الواضح أن الوسيط هو القيمة التي تقع في وسط التوزيع إذا كان عدد قيم المتغير فردياً أو هو متوسط القيمتين الوسطيتين إذا كان عدد القيم زوجياً . فمثلا للمجموعة .

٤ ٥ ٥ ٦ ٧ ٧ ٩ ١١ ١٢ ١٢ ١٤ ١٥ ١٥ ٤٠ ٥٠

التي عدد قيمها ١٥ يكون الوسيط هو العدد ١١ إذ أن هذا العدد يقسم المجموعة إلى قسمين متساويين في عدد القيم سبعة منها قبله وسبعة منها بعده .

أما بالنسبة للمجموعة

٤ ٥ ٥ ٦ ٧ ٧ ٩ ١١ ١٢ ١٢ ١٤ ١٥ ١٥ ٤٠

التي عدد قيمها ١٤ فيكون الوسيط هو العدد $\frac{٩+١١}{٢}$ = ١٠ إذ أنه يقسم المجموعة إلى قسمين متساويين في عدد القيم سبعة منها قبله وسبعة منها بعده مع ملاحظة أن الوسيط هنا ليس أحد قيم المجموعة المعاطاة .

بصفة عامة إذا رتبت قائمة من الأعداد حجمها ن لمتغير وثاب ترتيبا تصاعديا فإن :

(أ) ترتيب الوسيط هو $\frac{١}{٢}$ (ن + ١)

ففى القائمة الأولى لدينا ن = ١٥ .·. ترتيب الوسيط ٨ وعلى ذلك فالوسيط هو العدد الثامن فى القائمة أى العدد ١١ .

وفى القائمة الثانية لدينا ن = ١٤ .·. ترتيب الوسيط ٧,٥ وهذا العدد يعنى أن الوسيط هو متوسط العددين السابع والثامن فى القائمة أى العدد ١٠ .

(ب) يحسب ترتيب الربيع الأول من ترتيب الوسيط كالآتى :

ترتيب $ر_١$ هو $\frac{١}{٢}$ (ترتيب الوسيط بعد حذف الكسر إذا وجد + ١) .

(جـ)من التماثل يكون ترتيب الربيع الثالث هو نفس ترتيب الربيع الأول حين تقرأ قائمة الأعداد عكسيا من النهاية إلى البداية . .

ففى القائمة الأولى لدينا ن = ١٥ ، ترتيب الوسيط هو ٨

∴ ترتيب $ر_١$ هو $\frac{١}{٢}$ (٨ + ١) = ٤,٥

وهذا يعنى أن الربيع الأول هو متوسط العددين الرابع والخامس فى القائمة أى أن

$ر_١ = \frac{١}{٢}(٦+٧) = ٦,٥$

من التماثل

$ر_٣ = \frac{١}{٢}(١٥+١٤) = ١٤,٥$

وفى المجموعة الثانية لدينا ن = ١٤، ترتيب الوسيط هو ٧,٥

∴ ترتيب $ر_١$ هو $\frac{١}{٢}(٧+١) = ٤$

∴ $ر_١ = ٦$

من التماثل $ر_٣ = ٤٦$

(٢ -٤) الوصف العددي للتوزيعات التكرارية :

حين يكون لدينا توزيع لمتغير **عددي** وحين يكون لهذا التوزيع قمة واحدة كما يبدو مثلا من المنحني التكراري ، فإننا نستطيع وصفه موضوعياً من حيث عدة جوانب أهمها :

(أ) النزعة المركزية **Central Tendency**

(ب) التشتت **Dispersion**

(جـ) الإلتواء **Skewness**

(د) التفرطح **Kurtosis**

ففي أغلب التوزيعات ذات القمة الواحدة يتراكم عدد كبير من قيم المتغير حول قيمة معينة ويقل هذا التراكم تدريجياً على وجه العموم كلما ابتعدت القيم عن هذه القيمة من الجانبين . هذا التراكم أو التجمع حول قيمة مايسمى بالنزعة المركزية للتوزيع أو بالقيمة المتوسطة وتسمى القيمة التي يحدث حولها التجمع بمقياس النزعة المركزية . ويهمنا في دراسة التوزيعات الحصول على هذا المقياس . ولما كان هذا المقياس يختلف في تركيبه بحسب طبيعة البيانات والهدف من دراستها فقد وضعت عدة مقاييس للنزعة المركزية نختار منها مانرى أنه يلائم ما بأيدينا من توزيعات . ومن أشهر هذه المقاييس ما يلي :

الوسط الحسابي – الوسيط – المنوال – الوسط الهندسي – الوسط التوافقي .

وسنهتم هنا بصفة خاصة بالوسط الحسابي لأسباب عدة منها أنه أقوى مقاييس النزعة المركزية استجابة للمعالجة الرياضية ، وهو المقياس الذي نستخدمه عادة مالم يتضح لنا أنه لا يعبر تعبيراً صادقاً عن هذه النزعة كما هو الحال مثلا عندما يكون التوزيع مشتملا على قيم متطرفة تشذ عن بقية القيم .

كما يهمنا كذلك قياس تشتت التوزيع أي قياس مدى انحراف قيم المتغير بالنسبة إلى بعضها وبالنسبة إلى القيمة المتوسطة ، أو بمعنى معكوس ، مدى تجانس التوزيع . ومن أشهر مقاييس التشتت مايلي :

المدى – الانحراف الربيعي – الانحراف المتوسط – الانحراف المعياري – معامل الاختلاف .

وسنهتم هنا بصفة خاصة بالانحراف المعياري لأنه كمقياس للتشتت يتمشى مع الوسط الحسابي كمقياس للنزعة المركزية وهما يؤخذان معاً أو يتركان معاً .

(٢ - ٤ - ١) الوسط الحسابي والانحراف المعياري :

MEAN AND STNDARD DEVIATION

مثال (٢ - ٥) :

اعتبر الأعداد السبعة الآتية : ٥ ٨ ٧ ١٠ ١١ ٤ ٩ .

نعلم أن الوسط الحسابي لعدد من القيم هو مجموع هذه القيم مقسوماً على عددها وهذا تعريف عام للوسط الحسابي ويمكن صياغته رمزياً في حالتنا هذه كالآتي :

$$\text{الوسط الحسابي} = \overline{س} = \frac{\sum س_ر}{ن} \qquad (1)$$

حيث $س_ر$ تعبر عن قيم المتغير ، ن تعبر عن عدد هذه القيم .

ويعرف التباين Variance بأنه متوسط مجموع مربعات انحرافات قيم المتغير عن الوسط الحسابي وهذا ما نستطيع كتابته رمزياً كالآتي (على أساس أن التوزيع التكراري هو توزيع عينة) :

$$\text{التباين} = ع^2 = \frac{1}{ن - 1} \sum (س_ر - \overline{س})^2 \qquad (٢-أ)$$

يلاحظ أن هذا العدد يساوي صفراً إذا وإذا فقط تساوت جميع القيم $س_ر$ لأن كلا منها في هذه الحالة يساوي $\overline{س}$ ، وهناك صورة أخرى للتباين تشتق من هذه الصورة بعمليات جبرية بسيطة ، وهذه الصورة هي :

$$ع^2 = \frac{1}{ن - 1} \left[\sum س_ر^2 - \frac{(\sum س_ر)^2}{ن} \right] \qquad (٢ - ب)$$

وبالرغم من أن هاتين الصورتين متطابقتان رياضياً ، إلا أننا في حساب التباين نستخدم الصورة الثانية لأن الخطأ الذي ينتج فيها من تقريب الأعداد أقل من ذلك

الذي ينتج من الصورة الأولى ، كما أنها أكثر طواعية لحاسبات الجيب والحاسبات الإلكترونية .

أما **الانحراف المعياري** فيعرف بأنه الجذر التربيعي الموجب للتباين ويرمز له بالرمز ع . وهو عدد موجب دائماً بالتعريف .

من الصيغتين (١) ، (٢ –ب) نرى أن حساب الوسط الحسابي والتباين يعتمد على حساب المجموعين مجـ $س_ر$ ، مجـ $س_ر^٢$ وهذان المجموعان يمكن إيجادهما مباشرة من حاسبات الجيب أو من الجدول (٢ –١١) الآتي وهو يخص المثال (٢ –٥) .

الجدول (٢ –١١)

لإيجاد الوسط الحسابي والانحراف المعياري لمجموعة من القيم

$س_ر$	$س_ر^٢$
٥	٢٥
٨	٦٤
٧	٤٩
١٠	١٠٠
١١	١٢١
٤	١٦
٩	٨١
٥٤	٤٥٦

الوسط الحسابي $\bar{س} = \frac{مجـ س_ر}{ن} = \frac{٥٤}{٧} = ٧,٧١$ تقريباً

التباين $ع^٢ = \frac{١}{ن - ١} \left[مجـ س_ر^٢ - \frac{(مجـ س)^٢}{ن}\right] = \frac{١}{٦} \left[٤٥٦ - \frac{(٥٤)^٢}{٧}\right]$

$= ٦,٥٧١$ تقريباً

الانحراف المعياري $ع = \sqrt{٦,٥٧١} = ٢,٥٦$ تقريباً

مثال (٢ - ٦) :

أوجد الوسط الحسابي والإنحراف المعياري للتوزيع التكراري البسيط المعطى بالمثال (٢ - ١) السابق .

نظراً لأن كل قيمة $س_ر$ من قيم المتغير مكررة $ك_ر$ من المرات فإن مجموع القيم يكون $مجـ ك_ر س_ر$ وعلى ذلك فإن التعريف العام للوسط الحسابي يمكن صياغته في هذه الحالة كالآتي :

$$الوسط الحسابي\ \bar{س} = \frac{مجـ ك_ر س_ر}{ن} \qquad (٣)$$

$$حيث\ ن = مجـ ك_ر$$

وبالمثل نعرف التباين كالآتي :

$$التباين = ع^٢ = \frac{١}{ن - ١} مجـ ك_ر (س_ر - \bar{س})^٢ \qquad (٤ - أ)$$

$$أو \qquad = \frac{١}{ن - ١} \left[مجـ ك_ر س_ر^٢ - \frac{(مجـ ك_ر س_ر)^٢}{ن}\right] \qquad (٤ - ب)$$

يلاحظ أن التعريفين (١) ، (٢) ما هما إلا حالة خاصة من التعريفين (٣) ، (٤) تكون فيها جميع التكرارات ك ر مساوية للواحد .

لإيجاد المجموعين مجـ ك ر س ر ، مجـ ك ر س ر ٢ توطئة لحساب الوسط الحسابي والتباين نستخدم جدولاً كالجدول (٢ -١٢) الآتي :

الجدول (٢ -١٢)

لإيجاد الوسط الحسابي والانحراف المعياري لتوزيع تكراري بسيط

س ر	ك	ك ر س ر	ك ر س ر ٢
٧٧	٢	١٥٤	١١٨٥٨
٧٨	١	٧٨	٦٠٨٤
٧٩	٢	١٥٨	١٢٤٨٢
٨٠	٢	١٦٠	١٢٨٠٠
٨١	١	٨١	٦٥٦١
٨٢	٢	١٦٤	١٣٤٤٨
٨٣	٢	١٦٦	١٣٧٧٨
٨٤	٢	١٦٨	١٤١١٢
٨٥	٣	٢٥٥	٢١٦٧٥
٨٦	٥	٤٣٠	٣٦٩٨٠
٨٧	٣	٢٦١	٢٢٧٠٧
	٢٥	٢٠٧٥	١٧٢٤٨٥

$$\text{الوسط الحسابي} = \bar{س} = \frac{مج\ ك_ر\ س_ر}{ن} = \frac{٢٠٧٥}{٢٥} = ٨٣$$

$$\text{التباين} = ع^٢ = \frac{١}{ن - ١}\left[مج\ ك_ر\ س_ر^٢ - \frac{(مج\ ك_ر\ س_ر)^٢}{ن}\right]$$

$$= \frac{١}{٢٤}\left[١٧٧٢٤٨٥ - \frac{(٢٠٧٥)^٢}{٢٥}\right] = ١٠,٨٣$$

الانحراف المعياري = ع = ٣,٢٩١

الانحراف المعياري = موجب جذر التباين = $\sqrt{١٠,٨٣}$

مثال (٢ -٧) :

أوجد الوسط الحسابي والانحراف المعياري للتوزيع التكراري ذي الفئات المعطى بالمثال (٢ -٢) السابق .

في حالة التوزيع التكراري ذي الفئات نعتبر كما سبق الذكر أن جميع القيم الواقعة في فئة ما مساوية لمركز الفئة . وعلى ذلك يكون مجموع هذه القيم مج ك$_ر$ س$_ر$ حيث س$_ر$ هنا ترمز إلى مركز الفئة ، وتكون التعاريف (٣) ، (٤ -أ) ، (٤ -ب) صالحة للاستخدام هنا لإيجاد الوسط الحسابي والتباين مع فارق واحد وهو أن س$_ر$ ترمز إلى مراكز الفئات في حالة التوزيعات التكرارية ذوات الفئات ، بينما ترمز إلى قيم المتغير في حالة التوزيعات التكرارية البسيطة . ويجري الحساب كما في الجدول (٢ - ١٣) الآتي :

الجدول (٢ -١٣)

لإيجاد الوسط الحسابي والانحراف المعياري لتوزيع تكراري ذي فئات

الفئات	س ر	ك ر	ك ر س ر	ك ر $س_ر^2$
١٠,١٥ – ١٠,٤٥	١٠,٣	٣	٣٠,٩	٣١٨,٢٧
١٠,٤٥ – ١٠,٧٥	١٠,٦	٤	٤٢,٤	٤٤٩,٤٤
١٠,٧٥ – ١١,٠٥	١٠,٩	٤	٤٣,٦	٤٧٥,٢٤
١١,٠٥ – ١١,٣٥	١١,٢	٧	٧٨,٤	٨٧٨,٠٨
١١,٣٥ – ١١,٦٥	١١,٥	٥	٥٧,٥	٦٦١,٢٥
١١,٦٥ – ١١,٩٥	١١,٨	٩	١٠٦,٢	١٢٥٣,١٦
١١,٩٥ – ١٢,٢٥	١٢,١	٤	٤٨,٤	٥٨٥,٦٤
١٢,٢٥ – ١٢,٥٥	١٢,٤	٢	٢٤,٨	٣٠٧,٥٢
١٢,٥٥ – ١٢,٨٥	١٢,٧	٠	٠	٠
١٢,٨٥ – ١٣,١٥	١٣,٠	١	١٣,٠	١٦٩,٠٠
١٣,١٥ – ١٣,٤٥	١٣,٣	١	١٣,٣	١٧٦,٨٩
المجموع		٤٠	٤٥٨,٥	٥٢٧٤,٤٩

الوسط الحسابي لمحيط رأس الحمامة = $\bar{س} = \frac{٤٥٨,٥}{٤٠}$ = ١١,٤٦٢٥ مليمترا

التباين = $ع^2 = \frac{1}{39}\left[٥٢٧٤,٤٩ - \frac{(٤٥٨,٥)^2}{٤٠} \right]$

= ٠,٤٨٥٥

الانحراف المعياري = $\sqrt{٠,٤٨٥٥}$ = ٠,٦٩٦٨ مليمترا

ملاحظات :

(١) الأصل في تعريف التباين هو $\frac{1}{ن}$ مجـ كر (سر − $\bar{س}$)^2 ولكن حين نستخدم تباين عينة حجمها ن ووسطها الحسابي $\bar{س}$ لتقدير تباين مجتمع تقديرا غير متحيز نضع $\frac{1}{ن - 1}$ بدلا من $\frac{1}{ن}$ لأن تباين العينة يكون عادة أصغر من تباين المجتمع وهذا التعديل يجعل تباين العينة أكثر ملاءمة لتقدير تباين المجتمع ، وهناك من النظريات الرياضية ما يؤيد ذلك . وفيما عدا هذه الحالة نستخدم الصيغة الأصلية للتباين أي نحتفظ بالعدد ن .

(٢) لا تتغير قيمة التباين إذا طرح (أو أضيف) أي عدد من جميع قيم المتغير .

(٢ -٤ -٢) معامل الاختلاف :

يعرف معامل الاختلاف م . خ . لتوزيع ما كالآتي :

$$م . خ = \frac{ع}{\bar{س}} \times ١٠٠ \qquad (٥)$$

حيث ع هو الانحراف المعياري للتوزيع ، $\bar{س}$ وسطه الحسابي .

فمثلا ، للتوزيع الذي بالمثال (٢ -٦) :

$$م . خ = \frac{٣,٢٩١}{٨٣} \times ١٠٠ = ٣,٩٦٥$$

وللتوزيع الذي بالمثال (٢ -٧) :

$$م . خ . = \frac{٠,٦٩٦٨}{١١,٤٦٢٥} \times ١٠٠ = ٦,٠٧٩$$

إن معامل الاختلاف هو مقياس مطلق للتشتت وهو يستخدم لمقارنة تشتتات التوزيعات خاصة في الحالتين الآتيتين :

(أ) حين تختلف متوسطات التوزيعات اختلافاً كبيراً كما هو الحال عند مقارنة تشتت أطوال أذيال الأفيال وتشتت أطوال أذيال الفيران .

(ب) حين تختلف الوحدات التي تقاس بها المتغيرات كما هو الحال عند مقارنة تشتت أطوال مجتمع ما بأوزان هذا المجتمع .

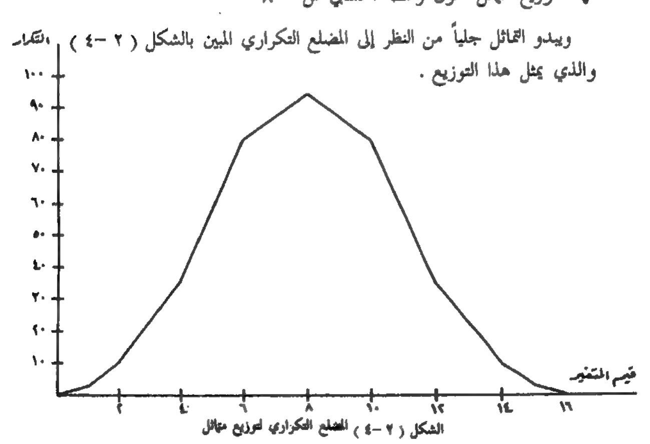
(٢ - ٤ - ٣) الالتواء : SKEWNESS

التواء توزيع ما يعني مدى بعده عن التماثل . ويكون التوزيع التكراري متماثلا إذا كانت التكرارات موزعة توزيعاً متماثلا حول الوسط الحسابي ، بمعنى أن تكون لقيم المتغير المتساوية البعد عن الوسط الحسابي نفس التكرارات . والتوزيع الآتي مثال لذلك :

س	١	٢	٤	٦	٨	١٠	١٢	١٤	١٥
ك	٣	١٠	٣٥	٨٠	٩٥	٨٠	٣٥	١٠	٣

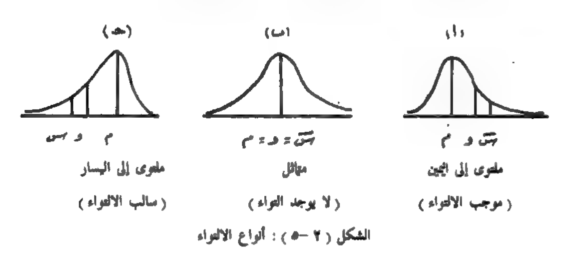
فهذا توزيع متماثل حول وسطه الحسابي $\bar{س} = ٨$

ويبدو التماثل جلياً من النظر إلى المضلع التكراري المبين بالشكل (٢ - ٤) والذي يمثل هذا التوزيع .



الشكل (٢ - ٤) المضلع التكراري لتوزيع متماثل

وتقسم المنحنيات التكرارية من حيث الالتواء إلى ثلاثة أنواع تتبين من الشكل (٢ –٥) الآتي حيث سَ ترمز إلى الوسط الحسابي ، و ترمز إلى الوسيط ، م ترمز إلى المنوال وهو القيمة الأكثر تكرارا في التوزيع .



الشكل (٢ –٥) : أنواع الالتواء

إذا كانت هذه الأشكال تمثل توزيعا لدرجات طلاب في امتحان ما فالشكل (أ) يشير إلى أن عددا كبيرا من الطلاب حصلوا على درجات أقل من المتوسط مما قد يعني أن مستوى الطلاب أقل من مستوى الامتحان . . . ، والشكل (ﺣ) يشير إلى عكس ذلك . وهناك حقيقة هامة مثلت بوضوح في هذه الأشكال نقدمها كما يلي :

إذا كان التوزيع ملتويا إلى اليمين فإن $\bar{س} >$ و $>$ م والعكس بالعكس .

وإذا كان التوزيع متماثلا فإن $\bar{س} =$ و $=$ م والعكس بالعكس

وإذا كان التوزيع ملتويا إلى اليسار فإن م $>$ و $> \bar{س}$ والعكس بالعكس .

ويقاس الالتواء عادة بأحد المقياسين الآتيين :

(١) مقياس الالتواء $= \frac{\text{٣ (الوسط الحسابي – الوسيط)}}{\text{الانحراف المعياري}} = \frac{\text{٣ } (\bar{س} - \text{و})}{\text{ع}}$ (٦)

وهذا المقياس مبني على الحقيقة سابقة الذكر .

(٢) معامل الالتواء = $\frac{١}{ن - ١}$ مج ك~ر~ (س~ر~ − س̄)^{٣}/ ع^{٣} (٧)

وهذا المقياس يتمشى مع الوسط الحسابي كمقياس للنزعة المركزية والانحراف المعياري كمقياس للتشتت .

من المهم أن نلاحظ مايلي :

(أ) كل من المقياسين (٦) و(٧) هو مقياس نسبي خالي من وحدات القياس وبالتالي يمكن استخدامه للمقارنة بين التواءات التوزيعات .

(ب) كل من المقياسين تقع قيمه بين العددين −٣ ، + ٣ .

(حـ) في كل من هذين المقياسين حين تكون القيمة موجبة نقول إن الالتواء موجب أو إنه التواء إلى اليمين ، وحين تكون سالبة نقول إن الالتواء سالب أو إنه التواء إلى اليسار . أما إذا كانت القيمة الناتجة صفرا فنقول إنه لا يوجد التواء أو إن التوزيع متماثل .

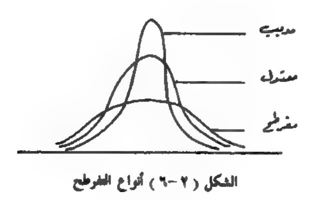
(٢ − ٤ − ٤) التفرطح : KURTOSIS (OR PEAKEDNESS)

تقسم المنحنيات التكرارية من حيث تفرطح قمتها إلى ثلاثة أنواع هي :

(أ) معتدلة (متوسطة التفرطح) . Mesokurtic (Normal)

(ب) مدببة . Leptokurtic

(ج) مفرطحة . Platykurtic



الشكل (٢ −٦) أنواع التفرطح

إن وصف المنحنيات بأنها مدببة أو مفرطحة يكون بالمقارنة مع المنحنيات المعتدلة التي سنتناولها بالدراسة في فصل قادم . وحين نقول إن المنحني مدبب فنحن نعني أن عدداً كبيراً من المفردات يتراكم بالقرب من الوسط الحسابي وعند الذيلين ولا يكون بالمواضع الأخرى إلا عدداً قليلا منها ، وذلك بالمقارنة مع المنحنيات المعتدلة . كذلك حين نقول إن المنحني مفرطح فنحن نعني أن عدداً قليلاً من المفردات يتراكم بالقرب من الوسط الحسابي وعند الذيلين ويكون هناك عدد كبير منها بالمواقع الأخرى ، وذلك بالمقارنة مع المنحنيات المعتدلة .

ويعرف المقياس الذي سنأخذه للتفرطح كالآتي وهو يتمشي مع الوسط الحسابي والتباين ومعامل الالتواء السابق تعريفها وجميعها من فصيلة تسمى بفصيلة العزوم :

$$\text{معامل التفرطح} = \frac{1}{\text{ن} - 1} \text{مج ك}_{\text{ر}} (\text{س}_{\text{ر}} - \bar{\text{س}})^4 / \text{ع}^4 \qquad (8)$$

حيث ع ترمز إلى الانحراف المعياري . وإذا وجد أن قيمة هذا المعامل في عينة ما قريبة من العدد ٣ قيل إن المنحني معتدل التفرطح ، وإذا زادت عن هذا العدد قيل إن المنحني مدبب ، وإذا قلت قيل إنه مفرطح .

ملاحظة :

إن الوسط الحسابي والانحراف المعياري ومعامل الالتواء ومعامل التفرطح ، بالإضافة إلى حجم التوزيع هي جل مانحتاج إليه في التحليل الوصفي للتوزيعات ذات القمة الواحدة ، وإذا كان التوزيع يمثل عينة لمجتمع ما فإن هذه القيم تتخذ أساساً لتقدير المعالم الإحصائية لهذا المجتمع باستخدام الطرق الإحصائية كما سنرى بعد . على أن هناك توزيعات لا تصلح هذه المقاييس لوصفها ويستلزم الأمر حينئذ اختيار مقاييس أخرى تناسب هذه التوزيعات . فمثلاً حين يكون التوزيع شديد الالتواء أو محتوياً على قيم متطرفة تشذ عن بقية القيم لا يكون الوسط الحسابي معبراً تعبيراً صادقاً عن النزعة المركزية ولا يكون الانحراف المعياري معبراً تعبيراً صادقاً عن التشتت ويتضح هذا من المثال الآتي :

مثال (٢ -٨) :

القيم الآتية هي أعمار ١٥ مريضاً بالسنوات دخلوا أحد أقسام إحدى المستشفيات في يوم ما ، وذلك بعد ترتيب هذه القيم تصاعدياً :

٣ ٣ ٤ ٥ ٦ ٧ ٨ ١٠ ١١ ١١ ١٢ ١٣ ١٤ ٥٠ ٦٠

نلاحظ أن هناك قيمتين تشذان عن بقية القيم وهما ٥٠ ، ٦٠ وإذا حسبنا الوسط الحسابي لهذه المجموعة نجده يساوي $\frac{٢١٧}{١٥}$ = ١٤,٥ سنة ولا يعقل أخذ هذه القيمة للتعبير عن متوسط أعمار المرضى فهي تزيد عن جميع قيم المجموعة المعطاة ماعدا قيمتين وفي الوقت ذاته تقل كثيراً عن هاتين القيمتين .

ويفضل في هذه الحال استخدام الوسيط كمقياس للنزعة المركزية لأنه لا يتأثر بالقيم المتطرفة . والوسيط هنا هو العدد ١٠ ومن الواضح أن هذا العدد يتوسط التوزيع وهو أصدق تعبيراً عن متوسط الأعمار من الوسط الحسابي .

في مثل هذه الحالات يستخدم مايسمى بنصف المدى الربيعي لقياس التشتت ومايسمى بمعامل الالتواء الربيعي لقياس الالتواء وهما مقياسان يتمشيان مع الوسيط ويعرفان بدلالة الربيعات كالآتي :

نصف المدى الربيعي = $\frac{1}{2}(ر_3 - ر_1)$ **Semi-interquartile range** (٩)

معامل الالتواء الربيعي = $\frac{(ر_3 - ر_2) - (ر_2 - ر_1)}{(ر_3 - ر_2) + (ر_2 - ر_1)}$ (١٠)

$$= \frac{ر_3 - 2\,ر_2 + ر_1}{ر_3 - ر_1}$$

ويلاحظ أن هذا المقياس للالتواء هو مقياس مطلق لا يتوقف على وحدات القياس وبالتالي يمكن استخدامه للمقارنة بين التواءات التوزيعات . وهو يساوي صفراً للتوزيعات المتماثلة حيث يقع الربيعان الأول والثالث على بعدين متساوين من الوسيط $ر_٢$. كما يلاحظ أن قيمة هذا المقياس تقع بين العددين −١ ، +١ وكلما ابتعدت قيمته عن الصفر من اليمين أو اليسار كلما دل ذلك على التواء التوزيع . كذلك :

$$\text{معامل الاختلاف الربيعي} = \frac{\text{نصف المدى الربيعي}}{\text{الوسيط}} \times ١٠٠ \qquad (١١)$$

وللمثال (٢ −٨) الأخير نجد − كما في البند (٢ −٣ −١) ومع ملاحظة أن المتغير متصل − مايلي :

$$ر_١ = ٥,٢٥ \text{ ، } ر_٢ = ١٠ \text{ ، } ر_٣ = ١٢,٧٥$$

$$\text{نصف المدى الربيعي} = \frac{١}{٢} (١٢,٧٥ - ٥,٢٥) = ٣,٧٥$$

$$\text{، معامل الالتواء الربيعي} = \frac{١٢,٧٥ - ٢٠ + ٥,٢٥}{١٢,٧٥ - ٥,٢٥} = -٠,٢٧$$

تمارين (٢ −١)

(١) احسب الوسط الحسابي والانحراف المعياري ومعامل الاختلاف للأعداد الآتية

١٠٥ ٩٨ ١٢٠ ٩٥ ١١٥ ١١٠ ١٠٠ ١٢٥ ٩٢ ١٣٠

(٢) رصدت أعمار عينة من ٢٧ شخصا بالسنوات المختلفة عند إصابتهم بمرض ما فوجدت كالآتي :

٣٩ ٥٠ ٢٦ ٤٥ ٤٧ ٧١ ٥١ ٣٣ ٤٠ ٤٠ ٥١ ٥٩ ٤٨ ٤٤ ٤٧ ٦١
٤١ ٥٧ ٣٦ ٥٥ ٦٣ ٦٦ ٤٢ ٥٤ ٤٧ ٥٣ ٥٤

أوجد مج س ، مج $س^2$ ومنها احسب الوسط الحسابي والانحراف المعياري للعمر عند الإصابة بذلك المرض . (لا داعي لتكوين توزيع تكراري) .

(٣) فحص ١٢٢ قرنا من قرون شجرة السرقم (الأبانوس الكاذب **laburnum**) فوجد مايلي :

عدد البذور في القرن	١	٢	٣	٤	٥	٦	٧	٨
عدد القرون	٢٧	٣٢	٢١	١٥	١٣	٩	٤	١

(أ) احسب الوسط الحسابي والانحراف المعياري لعدد البذور في القرن .

(ب) ارسم المدرج التكراري للتوزيع .

(٤) قيست أطوال ٢٥ عظمة فخذ نوع من الحشرات (م م $\times ١٠^{-١}$) فوجدت كما يلي :

٣,٨ ٣,٦ ٤,٣ ٣,٥ ٤,٣ ٣,٣ ٤,٣ ٤,٣ ٣,٩ ٤,٢ ٣,٩ ٣,٨ ٣,٩ ٤,٤
٣,٨ ٤,٧ ٣,٦ ٤,١ ٣,٦ ٤,٤ ٤,٥ ٣,٦ ٤,٤ ٤,١

(أولا) كون جدولا تكراريا ذا فئات طول فئته ٠,٣٠ ومثله بيانيا بمضلع تكراري .

(ثانيا) احسب كلا من الوسط الحسابي $\bar{س}$ والانحراف المعياري ع لطول عظمة الفخذ .

(ثالثا) ارسم منحني التكرارات المتجمعة المئوية ومنه احسب مايلي :

(أ) النسبة المئوية لعدد الحشرات التي تقل أطوالها عن ٤ م $م^{-١٠}$

(ب) النسبة المئوية لعدد الحشرات التي تقع أطوالها بين العددين $\bar{س} \pm ع$

(ج) الوسيط أي طول عظمة الفخذ الذي تقل عنه أطوال ٥٠٪ من الحشرات .

(٥) التوزيعات الثلاثة الآتية هي توزيعات درجات مجموعة من ٢٤ طالبا في ثلاثة اختبارات . ارسم المضلع التكراري لكل منها واذكر تعليقاً عن التواء كل توزيع .

التوزيع الأول $س_ر$: ١ ٢ ٣ ٤ ٥ ٦ ٧ ٨ ٩ ١٠

$ك_ر$: ٤ ٦ ٥ ٢ ٢ ١ ١ ١ ١ ١

التوزيع الثاني $س_ر$: ١ ٢ ٣ ٤ ٥ ٦ ٧ ٨ ٩ ١٠

$ك_ر$: ١ ١ ٢ ٣ ٥ ٥ ٣ ٢ ١ ١

التوزيع الثالث $س_ر$: ١ ٢ ٣ ٤ ٥ ٦ ٧ ٨ ٩ ١٠

$ك_ر$: ١ ١ ١ ١ ١ ٢ ٢ ٥ ٦ ٤

(٦) الأعداد الآتية هي الزيادة في الوزن بالكيلوجرامات لمجموعة من ١٣ بقرة بعد فترة من نظام غذائي معين :

٢,٤ ٢,٩ ٢,٤ ٢,٣ ٢,١ ٢,٥ ٢,١ ٢,٧ ٥ ٢,٩ ٢,٢ ٢,٨ ٢,٩

أوجد الوسيط ونصف المدى الربيعي ومعامل الالتواء الربيعي .

ملاحظة : استخدام الحاسبات :

تستطيع الحاسبات الالكترونية القيام بكل دقة وسرعة بالعمليات التي تتطلبها دراسة البيانات الاحصائية بدءا من تكوين الجداول والتوزيعات التكرارية من واقع البيانات الخام إلى حساب مقاييس النزعة المركزية والتشتت والالتواء والتفرطح .

(٢ - ٥) شكل الساق والورقة : **Stem - and - Leaf Diagram**

إن أسلوب الأشكال المسماه بأشكال الساق والورقة هو تنويع لأسلوب التوزيعات التكرارية ، فهو يؤدي إلى تجميع أو تكثيف البيانات في عدد مناسب من الأقسام مثله في ذلك مثل التوزيعات التكرارية ، إلا أنه يتميز عنها بأمرين رئيسيين أولهما أنه يحتفظ بفردية كل عنصر من عناصر البيانات وثانيهما أنه يسهل لنا التعرف على البيانات وتكوين فكرة عن توزيع المتغير الذي تعبر عنه هذه البيانات . ولهذا يعتبر هذا الأسلوب من الأدوات الأولية المفيدة في عملية التحليل الاستطلاعي للبيانات .

وإنشاء شكل ساق وورقة هو أمر غاية في السهولة ، ولا يتطلب إلا فصل أرقام كل عدد في البيانات إلى جزءين أحدهما يسمى ساق والآخر يسمى ورقة . والمعتاد أن تؤخذ الورقة على أنها الرقم الأخير في العدد أي الرقم الذي في أقصى يمينه ، أما الساق فهي بقية الأرقام . وتوضع الأرقام التي ترمز إلى السيقان رأسيا ثم توضع الأوراق المصاحبة لكل ساق أفقيا كما في الأمثلة الآتية .

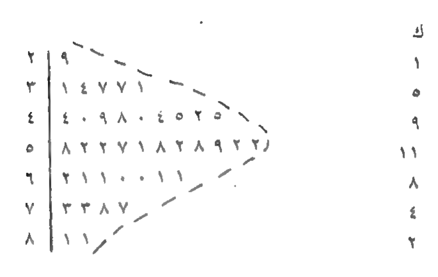
مثال (٢ -٩)

الأعداد الآتية هي الأعمار عند حدوث صدمة قلبية لعينة من أربعين مريضا . أنشىء شكل ساق وورقة واذكر ملاحظاتك عنه .

٣١	٥٨	٤٠	٣٧	٧٣	٧٨	٥٨	٤٢	٦٩	٨١
٦٢	٨١	٤٩	٤٨	٣١	٢٩	٦٠	٦١	٦٠	٥٢
٦١	٥٢	٣٤	٥٧	٧٣	٦١	٦٠	٥٢	٥٨	٥٢
٤٤	٥٢	٣٧	٥١	٤٠	٤٤	٤٥	٤٥	٥٩	٧٧

الحل :

في هذا المثال الورقة هي الرقم الأخير في العدد وهو رقم الآحاد أما الساق فهي رقم العشرات . فمثلا للعدد ٣١ الواحد هو الورقة والثلاثة هي الساق ، وللعدد ٦٢ الاثنين هي الورقة والستة هي الساق وهكذا .



الشكل (٢ –٧)

شكل ساق وورقة للأعمار التي حدثت عندها صدمات قلبية

لعينة من ٤٠ مريضا

لإنشاء الشكل نبدأ بتحديد السيقان وهي أرقام العشرات فنجد أنها تتراوح بين ٨،٢ . نرسم خطاً رأسيا ثم نكتب الأرقام الممثلة للسيقان على يساره مرتبة ترتيبا تصاعديا وتكون بذلك قد كتبنا جميع أرقام العشرات الممكنة ، ثم نمر على البيانات واحدا واحدا

لنكتب الرقم الممثل للورقة (أي رقم الآحاد) في كل منها في الصف الذي يناسبه أي على يمين الساق التي ترمز إلى رقم العشرات فيه . انظر الشكل (٢ –٧) . نلاحظ أنه في الصف الأول من الشكل يوجد عدد واحد فقط وهو يمثل العمر ٢٩ ، وفي الصف الثاني توجد خمسة أعداد هي الأعمار ٣١ ، ٣٤ ، ٣٧ ، ٣٧ ، ٣١ وهكذا . وإذا جمعنا الأعداد التي بالصفوف جميعها لوجدناها مساوية للعدد ٤٠ وهو حجم العينة .

ملاحظات :

١ – في بناء شكل الساق والورقة ينبغي أن نختار عددا مناسبا من السيقان ، وذلك لكي نستطيع الإفادة من الشكل ، والمعتاد ألا يقل هذا العدد عن خمسة وألا يزيد عن عشرين . هذا مع ملاحظة أن شكل الساق والورقة لا تكون له فائدة كبيرة إذا كان عدد البيانات كبيرا جدا أو كان المدى الذي يتغير فيه المتغير كبيرا .

٢ – يمكننا دائما استعادة البيانات الأصلية من الشكل المرسوم وذلك بضم الساق مع كل ورقة من أوراقه ، وهذا مانعنيه بقولنا أن الشكل يحتفظ بفردية البيانات .

٣ – لتسهيل التعرف على خصائص توزيع البيانات ، ندير الشكل بحيث يصبح الخط الرأسي أفقيا وتكون الأرقام الممثلة للسيقان أسفل هذا الخط ، ويساعدنا في ذلك أيضا أن ترسم خطا ناعما حول نهايات الأوراق ، ثم نحاول الإجابة عن تساؤلات كالآتية :

(أ) هل تميل البيانات إلى التجمع حول ساق أو سيقان معينة أم تتوزع على كل السيقان بشكل متعادل ؟

(ب) هل تتشتت البيانات تشتتا واسعا أم ضيقا ؟

(ج) هل هناك تماثل في توزيع البيانات ؟ هل تميل البيانات إلى التناقص تدريجيا نحو أحد طرفي التوزيع ؟ هل هناك مميزات خاصة يشير إليها المنحنى المرسوم حول نهايات الأوراق ؟

ففي المثال (٢ — ٩) نجد أن عددا كبيرا من البيانات (١١ عمرا) يتجمع حول الساق (٥) ونلاحظ أن الشكل يكاد يكون متماثلا حول هذه الساق ، كما نلاحظ أن الصدمات القلبية في هذه العينة يحدث أغلبها في الخمسينات ثم في الأربعينات والستينات ، وأن هذه الصدمات لا تحدث تقريبا قبل سن الثلاثين أو بعد سن الثمانين .

مثال (٢ —١٠)

ارسم شكل ساق وورقة للبيانات الآتية التي هي مشاهدات عن المتغير العشوائي الذي يعبر عن شدة الزلازل التي حدثت في أحد المناطق مقاسة بمقياس ريختر **Richter**. علق على الشكل .

١,٠ ٨,٣ ٣,١ ١,١ ٥,١ ١,٢ ١,٠ ٤,١ ١,١ ٤,٠
٢,٠ ١,٩ ٦,٣ ١,٤ ١,٣ ٣,٣ ٢,٢ ٢,٣ ٢,١ ٢,١
١,٤ ٢,٧ ٢,٤ ٣,٠ ٤,١ ٥,٠ ٢,٢ ١,٢ ٧,٧ ١,٥

الحل :

في أي عدد في هذه البيانات الورقة هي الرقم الذي في خانة الجزء من عشرة ، والساق هي رقم الآحاد . تتراوح السيقان بين ١ ، ٨ وإذن نكتب الأرقام ١ ، ٢ ، ... ، ٨ على يسار الخط الرأسي ثم نسجل الأوراق كل أمام الساق المناسبة فنحصل على الشكل (٢ —٨) الآتي حيث نلاحظ أن هناك ١١ عددا في الصف الأول هي : ١,٠ ، ١,٤ ، ١,٩ ، ١,١ ، ١,٤ ، ١,٣ ، ١,٢ ، ١,٠ ، ١,٢ ، ١,١ ، ١,٥ ، وثمانية أعداد في الصف الثاني هي : ٢,٠ ، ٢,٧ ، ٢,٤ ، ٢,٢ ، ٢,٢ ، ٢,٣ ، ٢,١ ، ٢,١ ، وهكذا .

الساق	الأوراق	التكرار
١	٠ ٤ ٩ ١ ٤ ٣ ٢ ٠ ٢ ١ ٥	١١
٢	٠ ٧ ٤ ٢ ٢ ٣ ١ ١	٨
٣	١ ٠ ٣	٣
٤	١ ١ ٠	٣
٥	١ ٠	٢
٦	٣	١
٧	٧	١
٨	٣	١
		٣٠

الشكل (٢ -٨)

شكل ساق وورقة لشدة الزلازل مقاسة بمقياس ريختر
في عينة مأخوذة من أحد المناطق

التعليق :

تتشتت شدة الزلازل بين القيمتين ١,٠ ، ٨,٣ غير أن البيانات تميل إلى التجمع حول القيم الصغرى وتقل تدريجيا في اتجاه القيم الكبرى (هناك التواء إلى اليمين) وهذا يعنى أن معظم الزلازل في هذه العينة كانت خفيفة . وإذا كانت هذه العينة تعبر عن المجتمع ككل ، فإن وقوع زلازل شديدة في هذه المنطقة يكون أمرا بعيد الاحتمال .

مثال (٢ -١١)

ارسم شكل ساق وورقة مع التعليق لبيانات المثال (٢ -٢) السابق عن أطوال محيطات رؤوس عينة من ٤٠ من الحمام المنزلي وهي :

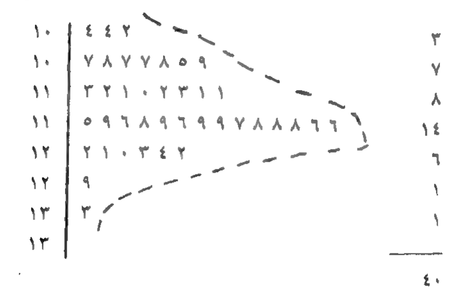
١١,٦ ١١,٨ ١١,٨ ١٢,٢ ١٢,٣ ١١,١ ١١,٩ ١١,٨ ١٢,٩ ١٢,٢

١١,١ ١٠,٥ ١١,٢ ١٣,٣ ١١,٩ ١١,٦ ١١,٢ ١١,٣ ١١,٥ ١٠,٧

١٢,١ ١١,٩ ١٠,٤ ١٠,٧ ١٠,٨ ١١,٠ ١١,٩ ١٠,٢ ١٠,٩ ١١,٦
١٠,٨ ١١,٦ ١٠,٤ ١٠,٧ ١٢,٠ ١٢,٤ ١١,٧ ١١,٨ ١١,٣ ١١,١

الحل :

في هذا المثال الورقة هي الرقم الذي بخانة الجزء من عشرة ، والساق هي العدد المكون من رقمي الآحاد والعشرات ، فمثلا للعدد ١٢,٢ الورقة هي ٢ والساق هي ١٢ . إذا استخدمنا الطريقة التي استخدمناها في المثالين السابقين نجد أن لدينا أربعة سيقان فقط هي ١٠ ، ١١ ، ١٢ ، ١٣ ويكون من الصعب تكوين فكرة عن التوزيع بهذا العدد القليل من السيقان . وللتغلب على هذه الصعوبة نأخذ كل رمز يرمز إلى ساق مرتين وبذلك نكون قد قسمنا البيانات إلى ٨ أقسام وهذا عدد مناسب ، على أن نكتب أمام الرمز في المرة الأولى الأوراق ٠ ، ١ ، ٢ ، ٣ ، ٤ ونكتب أمام الرمز في المرة الثانية الأوراق ٥ ، ٦ ، ٧ ، ٨ ، ٩ كما في الشكل (٢ -٩) الآتي :



الشكل (٢ -٩)

شكل ساق وورقة لأطوال محيطات رؤوس عينة من ٤٠ من الحمام المنزلي

التعليق :

تشتت أطوال محيطات رؤوس الحمام بين القيمتين ١٠,٢ ، ١٣,٣ من المليمترات ، إلا أن عددا كبيرا منها يتجمع حول الساق ١١ ويقل هذا التجمع تدريجيا من الناحيتين بمقادير متوازنة تكاد تجعل الشكل متماثلا .

تمارين (٢ - ٢)

١ - الآتي هي أعداد النقط التي حاز عليها ٤٠ لاعبا في فرق كرة القدم بإحدى المدارس الثانوية . انشيء شكل ساق وورقة مستخدما السيقان ٠ ، ١ ، ٢ ، ٣ ، ٤ ، ٥ ، ٦ ، ٧ ثم علق على الشكل .

٠	٥٧	٠	١٨	١	٠	٤٥	٥٩	٢	٠
٣	٣	٩	٤٢	٠	١	٤	٧	٦٠	٢
٤	١٧	٣٥	٣٧	٦	٣	٨	٤٦	٣	٤٨
٢	١٨	٠	٢	٢١	٦	٢١	٢	٧٥	٧٥

٢ - أجريت دراسة لمدى تأثير التدخين على نمط النوم . المتغير العشوائي س الذي يدرس هو الزمن بالدقائق الذي يمضي حتى ينام الشخص ، وقد وجدت البيانات الآتية في عينتين عشوائيتين إحداهما من المدخنين والأخرى من غير المدخنين . المطلوب رسم شكل ساق وورقة لكل عينة باستخدام الأعداد من ١٥ إلى ٢٥ كسيقان ثم بيان ما إذا كان هناك فرق بين توزيع المتغير س في العينتين .

المدخنون						غير المدخنين					
٢٤,٨	١٦,٠	٢١,٣	١٧,٧	٢٠,٥	١٥,١	١٧,٦	١٨,٣	١٥,١	١٨,١	١٩,٧	١٧,٢
٢٥,٢	١٥,٩	٢٢,١	١٨,١	٢١,٢	١٦,٨	٢٠,١	٢٤,٩	٢٣,٦	١٩,٨	١٩,٩	١٦,٢
٢٥,٠	١٨,٣	٢٥,٢	١٩,٤	٢٢,٤	٢٢,٨	٢١,٤	٢٥,٠	٢٤,١	٢٠,٠	٢٢,٦	١٩,٨
١٦,٣	٢١,٦	٢٤,١	١٥,٠	٢٤,١	٢٥,٨	٢٠,٢	٢٣,٣	٢٠,٦	٢٢,١	١٨,٩	٢١,٢
١٧,٩	٢٣,٨	١٨,٠	١٥,٢	٢٥,٧	٢٤,٣	٢١.٣	١٧,٥	٢٠,١	٢٣,٠	١٦,٩	٢١,١
١٩,٩	٢٤,٩	١٧,٢	١٦,١	٢٥,١	٢٣,٢	٢٠,٧	٢٠,٤	٢٠,٥	٢١,١	٢٢,١	٢١,٨
٢٥,١	٢٣,٠	٢٣,١	١٩,٩	١٥,٣	١٥,٧	١٧,٤	٩١,٣	٢٢,٤	١٩,٢	١٨,٨	١٩,٥

٣ – في تجربة نفسية عن التعلم استخدم ٤٠ فأرا قسموا عشوائيا إلى قسمين متساويين في العدد وأتيح لكل فأر أن يجري في متاهة وسجل الوقت الذي يستغرقه في اتمامها بالثواني . دربت واحدة فقط من المجموعتين على الجري في المتاهة ، ثم أتيح لكل فأر أن يجري في المتاهة مرة ثانية وسجل الوقت الذي يستغرقه في هذه المرة الثانية . المتغير الذي يدرس هو الفرق في الوقت بين المرتين (الوقت في المرة الأولى – الوقت في المرة الثانية) . دونت هذه الفروق في الجدول الآتي :

الفئران المدربة				الفئران غير المدربة			
٤,٩	٤,١	٣,٢	٤,٠	−٢,٥	−١,١	−٢,٢	−٢,١
٤,٢	٤,٣	٣,٧	٤,٢	−٠,٦	−٢,٤	٢,٠	−١,٢
٤,٩	٣,٥	٣,٦	٤,٤	−٢,٧	−٠,٢	−١,٣	١,٣
٥,٠	٤,٧	٤,٥	٥,١	٢,١	٢,٢	٠,٩	١,٤
٥,٥	٥,٢	٤,٦	٥,٦	٢,٦	١,١	٢,١	١,٨

انشىء شكل ساق وورقة لكل من المجموعتين (استخدم السيقان ٣ ، ٣ ، ٤ ، ٤ ، ٥ ، ٥ للفئران المدربة؟ −٢ ، −١ ، −٠ ، ١،٠ ، ٢ للفئران غير المدربة) ثم حاول الاجابة عما يأتي :

(أ) متى تنتج القيم الموجبة وماذا تعنى هذه القيم ؟

(ب) من الواضح أن متوسط الفروق في عينة الفئران المدربة (وجميعها موجبة) أكبر منه في عينة الفئران غير المدربة ، فهل هذا يعنى أن الفئران تتعلم من التدريب ؟

(جـ) قارن بين التوزيعين مستخدما شكلى الساق والورقة .

الفصل الثالث

بعض نماذج الاحتمال

SOME PROBABILITY MODELS

كما أشرنا فى نهاية البند (١ – ٣) ، نصف المتغير س بأنه **متغير عشوائي** إذا كانت قيمه أعداداً حقيقية يتأثر قياسها بعوامل عشوائية ويكون ظهورها مصحوباً باحتمالات محددة . ولكل متغير عشوائي توزيع يربط بين قيمه واحتمالاتها يسمى بتوزيع الاحتمال لهذا المتغير . وتنقسم توزيعات الاحتمال للمتغيرات العشوائية إلى نوعين هما توزيعات الاحتمال الوثابة وتوزيعات الاحتمال المتصلة بحسب كون المتغير من النوع الوثاب أو من النوع المتصل .

(٣ – ١) توزيعات الاحتمال الوثابة

DISCRETE PROBABILITY DISTRIBUTIONS

ليكن س متغيراً حقيقياً وثاباً يأخذ القيم

$س_١$ ، $س_٢$ ، $س_٣$ ، ... باحتمالات قدرها

$ل_١$ ، $ل_٢$ ، $ل_٣$ ، ... حيث $\sum ل_ر = ١$

إن تجمع قيم س مع احتمالاتها المناظرة في مجموعة من الأزواج المرتبة

{ ($س_١$ ، $ل_١$) ، ($س_٢$ ، $ل_٢$) ، ($س_٣$ ، $ل_٣$) ، ... } .

يسمى بتوزيع الاحتمال للمتغير العشوائي س .

وفي كثير من الأحيان يمكن أن نجد داله غير سالبة د حيث :

$$د_ر (س) = ل (س = س) \qquad (1)$$

أى حيث تكون قيمة الدالة د عند س مساوية لاحتمال أن يأخذ المتغير س القيمة س . وتسمى هذه الدالة حينئذ **بدالة كتلة الاحتمال للمتغير** س probability mass function

مثال (٣ – ١) :

التوزيع الآتي هو توزيع الاحتمال لعدد مرات إصابة الهدف لجندى ما يطلق بندقيته على هدف ثابت ٥ مرات .

عدد مرات إصابة الهدف س :	صفر	١	٢	٣	٤	٥
احتمال هذا العدد ل :	صفر	٠,٤	٠,٣	٠,٢	٠,١	صفر

لاحظ أن مجـ ل = ٠ + ٠,٤ + ٠,٣ + ٠,٢ + ٠,١ +٠ = ١

ويمكن هنا التعبير عن الاحتمالات بواسطة الدالة د المعرفة كالآتي :

د (س) = $\frac{٥ - س}{١٠}$ حيث س = ١ ، ٢ ، ٣ ، ٤ ، ٥

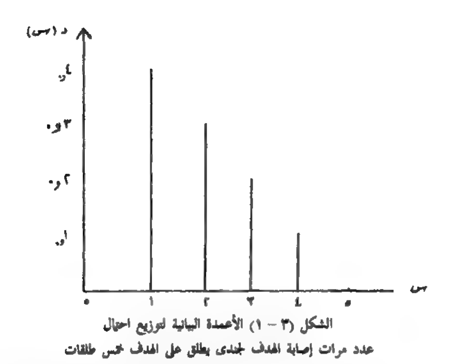
، صفر فيما عدا ذلك

إذ أنه بوضع س = ١ ، ٢ ، ٣ ، ٤ ، ٥ نحصل على الاحتمالات ٠,٤ ، ٠,٣ ، ٠,٢ ، ٠,١ ، صفر .

ونمثل هذه الدالة أو هذا التوزيع بيانيا كما في الشكل (٣ – ١) الآتي .

(٣–١–١) الوسط الحسابي والتباين : MEAN AND VARIANCE

إن توزيعات الاحتمال الوثابة تشبه التوزيعات التكرارية ، إلا أن الاحتمالات ل ر تحل محل التكرارات ك ر ، كما أن حجم التوزيع هو الواحد الصحيح دائماً . وهذا الواحد يشير إلى أن هناك احتمالا قدره الواحد الصحيح موزعاً على القيم المختلفة للمتغير ولذلك سمى التوزيع بتوزيع الاحتمال .



الشكل (٣ – ١) الأعمدة البيانية لتوزيع احتمال
عدد مرات إصابة الهدف لجندى يطلق على الهدف خمس طلقات

ومن هنا كان تعريفا الوسط الحسابي والتباين لتوزيع احتمال – وسنرمز لهما بالرمزين μ ، σ^2 ، يشبهان تعريفي الوسط الحسابي والتباين للتوزيع التكرارى فهما يعرفان كالآتي :

الوسط الحسابي = μ = مجـ لر سر (٢)

التباين = σ^2 = مجـ لر (سر – μ)^2

أو = مجـ لر س$^2_ر$ – μ^2 (٣)

أما الانحراف المعيارى فهو بالطبع الجذر التربيعى للتباين .

ففي المثال (٣ – ١) السابق نجد أن :

μ = مجـ لر سر

= ٠×٠ + ٠,٤ × ١ + ٠,٣ × ٢ +٠,٢×٣+ ٠,١ × ٤ + ٠ × ٥ =٢

، σ^2 = مجـ لر س$^2_ر$ – μ ٢

= (٠×٠+٠,٤×١+٠,٣×٤+٠,٢×٩+٠,١×١٦+٠×٢٥) –٤

= ٥–٤ = ١

(٣ – ٢) نماذج الاحتمال : PROBABILITY MODELS

نموذج الاحتمال لمتغير عشوائي س هو توزيع احتمال ذو صيغة رياضية محددة يفترض أنها تعكس سلوك المتغير س . ويعبر عن الاحتمالات من هذا النموذج بدلالة واحد أو أكثر من أدلة مجهولة تتوقف على خواص المجتمع وطريقة المعاينة منه . ويبني كل نموذج احتمال على افتراضات خاصة تصور تصويراً مناسباً الميكانيكية العشوائية التي تسبب الاختلافات في مشاهداتنا عن المتغير .

وسنتناول فيما يلي أربعة من أشهر نماذج الاحتمال للمتغيرات العشوائية الوثابة تعرف بتوزيع ذى الحدين وتوزيع بواسون وتوزيع باسكال والتوزيع الهندسي .

(٣ – ٣) توزيع ذى الحدين : THE BINOMIAL DISTRIBUTION

في كثير من الأحيان يكون اهتمامنا منصباً على وجود أو عدم وجود خاصة ما في وحدات أو عناصر مجتمع ما . ولذلك ننظر إلى المجتمع على أنه مقسم إلى قسمين منفصلين بحسب هذه الخاصة . فمثلا قد نقسم مجتمعاً من الطلاب بحسب خاصة القومية إلى القسمين : عربي / غير عربي أو بحسب الجنس إلى : ذكر/ أنثي أو إلى القسمين : يدخن / لا يدخن أو إلى القسمين : يلبس نظارة / لا يلبس نظارة أو إلى أى قسمين منفصلين متكاملين .

في مثل هذه الحال تتركز الصفة الأساسية للمجتمع في دليل أو بارامتر واحد هو نسبة أى من القسمين – وليكن القسم الأول – إلى المجتمع كله ، وسنرمز إلى هذه النسبة بالرمز ح . فمثلا قد تكون ح نسبة الطلاب العرب في المجتمع ، وهنا تكون نسبة الطلاب غير العرب هى ١ – ح = ك مثلا . وهذا يعني أننا إذا سحبنا عشوائياً عنصراً من المجتمع فإن ح تعبر عن احتمال أن يكون هذا العنصر من القسم الأول (طالب عربي مثلا) كما أن ك تعبر عن احتمال أن يكون العنصر من القسم الثاني (طالب غير عربي) .

سنسمى ظهور عنصر من القسم الأول **نجاحاً** للخاصة أو الحدث الذى ندرسه وظهور عنصر من القسم الثاني **فشلا** للحدث ، أى سنعتبر أن لدينا حدثاً واحداً — مثلا ظهور طالب عربي — إما أن يقع أو لا يقع .

إن هدفنا الأساسي من هذه الدراسة يتلخص فيما يلى : نفرض أننا سحبنا من المجتمع عينة عشوائية حجمها ن = ١٠ مثلا . هناك ١١ حالة ، إذ يمكن أن تكون هذه العينة خالية من أى عنصر من القسم الأول كما يمكن أن تشتمل على عنصر واحد فقط من هذا القسم أو تشتمل على عنصرين أو ثلاثة أو أربعة أو ... عشرة . أى أن عدد مرات نجاح الحدث يمكن أن يكون ٠ ، ١ ، ٢ ، ٣ ، ... ، ١٠ . والسؤال الذى نستهدف الإجابة عنه هو : ما احتمال كل من هذه الحالات ؟ وبمعني آخر إذا اعتبرنا أن لدينا متغيراً عشوائياً س يعبر عن **عدد مرات نجاح الحدث** فما توزيع الاحتمال لهذا المتغير ؟

إن الإجابة عن هذا السؤال تتوقف على ما نضعه من افتراضات نرى أنها مناسبة لما يهمنا من أوضاع ويمكن تحققها في عملية التجريب . وسوف نتبني هنا الافتراضات أو الشروط الآتية :

(١) عشوائية العينة :

سنفترض أن المعاينة (أى سحب العناصر من المجتمع) عشوائية .

(٢) ثبات الدليل ح :

سنفترض أن احتمال نجاح الحدث هو عدد **ثابت** ح طوال عملية سحب العينة . ولكى يتحقق هذا الفرض سنعتبر أن حجم المجتمع كبيراً بالنسبة لحجم العينة وبالتالى فإنه حتي إذا كانت المعاينة بغير إرجاع يكون التغير الذى يحدث في قيمة ح تغيراً طفيفاً يمكن التجاوز عنه ، كما سنفترض أيضاً أن قيمة ح لا تختلف من عينة إلى أخرى .

(٣) استقلال الأحداث :

سنفترض أن نجاح (أو فشل) الحدث في أى سحبة مستقل عما نتج من نجاح أو فشل في السحبات السابقة ، أى أن ما تسفر عنه أى سحبة لا يتأثر بأى حال بما نتج في السحبات الأخرى . كما سنفترض أن عدد مرات نجاح الحدث في عينة ما مستقل عن عدد مرات نجاحه في أى عينة أخرى .

تحت هذه الشروط وباستخدام قاعدة الضرب للأحداث المستقلة وقاعدة الجمع للأحداث المتنافية – انظر البند (١ – ٧) عن توافقات الاحتمال – نستطيع أن نثبت رياضياً أنه إذا كان المتغير س يعبر عن عدد مرات نجاح الحدث فإن دالة كتلة احتماله تأخذ الصورة الآتية :

$$\text{د}(\text{س}) = {}^{\text{ن}}\text{ق}_{\text{س}}\ \text{ح}^{\text{س}}\ \text{ك}^{\text{ن}-\text{س}} \quad \text{حيث س} = ٠ ، ١ ، \ldots ، \text{ن} \qquad (٤)$$

وحيث ١ > ح > . ، ك = ١ – ح ، ن هو حجم العينة .

ويسمى توزيع الاحتمال حينئذ بتوزيع ذى الحدين .

كما نستطيع أن نثبت رياضياً أن

$$\text{الوسط الحسابي للتوزيع} = \text{ن ح} \qquad (٥)$$

$$\text{تباين التوزيع} = \text{ن ح ك} \qquad (٦)$$

وجدير بالملاحظة أن الدالة (٤) تتوقف على اثنين من الأدلة هما ن ، ح ومعرفة هذين الدليلين تحدد التوزيع تحديداً تاماً . ولذلك سنرمز لتوزيع ذى الحدين بالرمز حد (ﻫ ، ح) .

ملاحظة :

يمكن ايجاد التوافقات ${}^{\text{ن}}\text{ق}_{\text{س}}$ بالطريقة الحسابية المعتادة أو باستخدام مثلث

باسكال أو باستخدام جداول جاهزة حسبت فيها هذه التوافقات – انظر الجدول (٢) في ملحق هذا الكتاب وهو يعطى التوافقات $\binom{ن}{س}$ لبعض قيم ن ، س (ن = ١ ، ٢ ، ٣ ، ... ، ٢٠ و س = ٠ ، ١ ، ٢ ، ... ، ٢٠) ، كما يمكن إيجاد قيم الاحتمالات (٤) من الجدول (٣) لبعض قيم ن ، ح (ن = ٢ ، ٣ ، ... ، ١٥ و ح = ٠,٠٥ ، ٠,١ ، ٠,٢ ، ٠,٣ ، ٠,٤ ، ٠,٥ ، ٠,٦ ، ٠,٧ ، ٠,٨ ، ٠,٩ ، ٠,٩٥) .

مثال (٣ – ٢) :

افرض أن احتمال ولادة مولود ذكر هو ٠,٥ واعتبر العائلات التي أنجبت ٤ أطفال .
(أ) أوجد توزيع احتمال المتغير سـ الذى يعبر عن عدد الذكور في العائلات ذوات الأربعة الأطفال .
(ب) إذا أخذنا عشوائياً ٢٠٠٠ عائلة من هذا النوع فما العدد الذى نتوقعه للعائلات التي يكون بها ولدين على الأقل ؟

الحـــل :

لدينا مجتمع من الولادات مقسم إلى القسمين ذكر / أنثي ، واحتمال وقوع أو نجاح الحدث « المولود ذكر » هو عدد **ثابت** ح = $\frac{1}{2}$. إن المتغير سـ يعبر هنا عن عدد الأولاد (الذكور) في العائلات ذوات الأربعة الأطفال أى عدد المرات التي تحدث فيها ولادة مولود ذكر في هذا النوع من العائلات وإذن فالمتغير سـ لا يأخذ إلا القيم ٠ ، ١ ، ٢ ، ٣ ، ٤ . إن كل عائلة تعتبر عينة **عشوائية** حجمها ن = ٤ . فإذا فرضنا أن إنجاب مولود ذكر (أو أنثي) في أى ولادة **مستقل** عما ينتج في الولادات الأخرى تكون الافتراضات الثلاثة لتوزيع ذى الحدين متوفرة ويكون للمتغير سـ توزيع ذى الحدين دليلاه ن = ٤ ، ح = $\frac{1}{2}$ وبالتالى تكون دالة كتلة احتماله :

$$د(س) = ق^{٤}_{س} \left(\frac{١}{٢}\right)^{س} \left(\frac{١}{٢}\right)^{٤-س} = \frac{١}{١٦} ق^{٤}_{س}$$

حيث س = ٠ ، ١ ، ٢ ، ٣ ، ٤

(أ) توزيع الاحتمال المطلوب هو التوزيع المبين في العمودين الأول والثاني من الجدول (٣ – ١) الآتي :

الجدول (٣ – ١)

توزيع الاحتمال لعدد الذكور في العائلات ذوات الأربعة الأطفال – ح = ٠,٥

عدد الذكور في العائلة	احتمال هذا العدد	العدد المتوقع من العائلات
س	د (س)	٢٠٠٠ × د (س)
٠	$\frac{١}{١٦} ق^{٤}_{٠} = \frac{١}{١٦}$	١٢٥
١	$\frac{١}{١٦} ق^{٤}_{١} = \frac{٤}{١٦}$	٥٠٠
٢	$\frac{١}{١٦} ق^{٤}_{٢} = \frac{٦}{١٦}$	٧٥٠
٣	$\frac{١}{١٦} ق^{٤}_{٣} = \frac{٤}{١٦}$	٥٠٠
٤	$\frac{١}{١٦} ق^{٤}_{٤} = \frac{١}{١٦}$	١٢٥
	١,٠٠٠	٢٠٠٠

(ب) أما العمود الثالث فيعطي الأعداد المتوقعة من العائلات التي بها ٠ ، ١ ، ٢ ، ٣ ، ٤ أولاد من بين الـ ٢٠٠٠ عائلة ، ومن هذا العمود ينتج أن العدد المتوقع للعائلات التي بها ولدين على الأقل .

= العدد المتوقع للعائلات التي بها ولدين أو ثلاثة أو أربعة .

= ٧٥٠ + ٥٠٠ + ١٢٥ = ١٣٧٥ عائلة .

مثال (٣ – ٣):

حسب نظرية مندل للخواص الوراثية ، حين يهجن نوع من النباتات حمراء

الزهور مع نوع ذى قرابة من نباتات بيضاء الزهور تنتج خلفة ٢٥٪ منها حمراء الزهور . نفرض أننا سنقوم بتهجين ٥ أزواج من هذه النباتات نختارها عشوائياً فما احتمال أن يكون من بين خمسة السلالات الناتجة :

(أ) لا توجد نباتات حمراء الزهور .

(ب) يوجد على الأقل ٤ نباتات حمراء الزهور ؟

الحل :

لدينا مجتمع من السلالات مقسم إلى القسمين : حمراء الزهور / بيضاء الزهور . واحتمال الحدث « حمراء الزهور » هو عدد ثابت ح = ٠,٢٥ ومن الطبيعى أن نعتبر أن النواتج مستقلة لأن التهجين يتم بين أزواج مختلفة من النباتات . وعلى ذلك تكون الشروط الثلاثة متوفرة ويكون لدينا متغير عشوائي سـ يعبر عن عدد النباتات حمراء الزهور في العينات العشوائية ذوات الحجم ن = ٥ وهو متغير له توزيع ذى الحدين دليلاه ٥ ، ٢٥ ,٠ ودالة كتلة احتماله

د(س) = ^{٥}ق$_{س}$ (٠,٢٥)^{س} (٠,٧٥)$^{٥-س}$ حيث س = ٠ ، ١ ، ٢ ، ٣ ، ٤ ، ٥

(أ)احتمال عدم وجود نباتات حمراء الزهور .

ل (س = ٠) = د (٠) = ^{٥}ق$_{٠}$ (٠,٢٥)^{٠} (٠,٧٥)^{٥} = (٠,٧٥)^{٥}

= ٠,٢٣٤

(ب) احتمال وجود ٤ نباتات حمراء الزهور على الأقل .

= ل (سـ ≥ ٤) = د (٤) + د (٥)   حدثان متنافيان

= ^{٥}ق$_{٤}$ (٠,٢٥)^{٤} (٠,٧٥) + (٠,٢٥)^{٥}

= ٠,٠١٦

(٣ – ٣ – ١) تقدير الدليل ح :

فى توزيع ذى الحدين ، إذا كانت قيمة البارامتر ح مجهولة ، يمكن تقديرها تجريبياً من عينات بحيث تتوفر الشروط الثلاثة سالفة الذكر . فإذا حصلنا على التوزيع التكرارى لعينة حجمها ن ووجدنا أن وسطها الحسابي $\bar{س}$ فإننا نأخذ هذا الوسط كتقدير للوسط الحسابي لتوزيع ذى الحدين وهو كما نعلم يساوى ن ح وبالتالى نقدر البارامترح بالعدد رحيث :

$$ر = \frac{\bar{س}}{ن} \qquad (٧)$$

مثال (٣ – ٤) :

لتقدير نسبة الحصي الجرانيتية إلى مجتمع الحصي على أحد الشواطىء أخذت من هذا الشاطىء ١٠٠ عينة عشوائية بكل منها ثلاث حصوات وحسب عدد الحصوات الجرانيتية بكل عينة فنتج التوزيع التكرارى الآتي :

عدد الحصوات الجرانيتية في العينة	$س_ر$:	٠	١	٢	٣
عدد العينات	$ك_ر$:	٥٨	٣٣	٧	٢

الوسط الحسابي لهذا التوزيع التكرارى = $\bar{س} = \frac{١}{ن}$ مجـ $ك_ر\, س_ر$

$= \frac{١}{١٠٠} (٠ + ٣٣ + ١٤ + ٦) = ٠,٥٣$

إذن تقدير البارامتر ح من العينة هو $ر = \frac{\bar{س}}{ن} = \frac{٠,٥٣}{٣} = ٠,١٧٧$ تقريباً

هذا على فرض أن توزيع عدد الحصي الجرانيتية هو توزيع ذى الحدين دليلاه ن ، ح حيث ن = ٣ ، ولنا أن نقول حينئذ أن هناك حوالى ١٧,٧٪ من الحصي الجرانيتية في مجتمع الحصي الذى على ذلك الشاطىء . ويمكننا اختبار مدى صواب هذا القول كما في البند التالى .

ويلاحظ أنه يمكن أيضاً تقدير النسبة ح من عينة عشوائية واحدة بشرط أن تكون كبيرة الحجم وذلك بأخذ التكرار النسبي لعدد الحصى الجرانيتية التي ظهرت في العينة . ففى هذا المثال لدينا ٣٠٠ حصوة منها ٥٣ حصوة جرانيتية (٠ + ٣٣ + ١٤ + ٦) وعلى ذلك فالتكرار النسبى للحصى الجرانيتية هو ٥٣/٣٠٠ = ٠,١٧٧

(٣ - ٣ - ٢) اختبار ما إذا كان الدليل ح له قيمة معينة . توفيق توزيع ذى الحدين لتوزيع تكرارى معلوم .

نفرض أن قائلا ذكر أن الدليل ح لتوزيع ذى الحدين لمتغير ما له قيمة معينه أ مثلا ونريد اختبار هذا القول . لتحقيق هذا الغرض نتبع الخطوات الثلاث الآتية :

(أ) نختار عينات عشوائية من حجم معين ن ونحسب عدد مرات وقوع الحدث في كل منها أى نحسب العدد ر (حيث ر = ٠ ، ١ ، ٢ ، .... ، ن) في كل منها . نجمع هذه النتائج في توزيع تكرارى لنحصل على التكرارات ك٠ ، ك١ ، ك٢ ، ... ، كن المناظرة للأعداد ٠ ، ١ ، ٢ ، ... ، ن . إن هذه التكرارات نرمز لها بالرمز كر ونسميها بالتكرارات المشاهدة . إن التوزيع الناتج يكون على نمط التوزيع الوارد بالمثال (٣ - ٤) .

(ب) إذا كان القول أو الفرض ح = أ صحيحا فإن توزيع الاحتمال للمتغير ذى الحدين يكون دليلاه ن ، أ معروفين ونستطيع إيجاد هذا التوزيع والحصول على الاحتمالات ل٠ ، ل١ ، ... ، لن المناظرة للأعداد ٠ ، ١ ، ٢ ، ... ، ن . بضرب كل من هذه الاحتمالات (التكرارات النسبية النظرية أو المتوقعة) في عدد العينات المأخوذة نحصل على الأعداد ق٠ ، ق١ ، ق٢ ، ... ، قن . إن هذه الأعداد نرمز لها بالرمز قر ونسميها بالتكرارات النظرية أو المتوقعة .

(حـ) نقارن بين التكرارات النظرية قر والتكرارات المشاهدة كر المناظرة لها فإذا كانت المطابقة حسنة أى كانت أزواج التكرارات قريبة من بعضها بدرجة

معقولة بحيث لا يكون بينها فروق كبيرة جاز لنا قبول الفرض أن ح = أ وإلا نرفضه .

إن مثل هذا الاختبار يدخل في موضوع اختبارات الفروض الذى سنتناوله في فصل لاحق حيث ستتعرف على أدوات تمكننا من الحكم حكماً موضوعياً على مدى صغر أو كبر مثل هذه الفروق وبالتالى من الحكم على صواب أو خطأ ذلك الفرض .

مثال (٣ – ٥) :

لاختبار الفرض القائل أن إنجاب الذكور وإنجاب الإناث متساويا الاحتمال ، اختيرت عشوائياً ٣٢٠ عائلة بكل منها ٤ أطفال فنتج التوزيع التكرارى الآتي :

عدد الأطفال الذكور في العائلة س ر	:	٠	١	٢	٣	٤
عدد العائلات ك ر	:	١٣	٤٣	٩٢	١١٢	٦٠

هل هذه البيانات تدعم الفرض المذكور أو تنفيه ؟

الحل :

على فرض أن إنجاب الذكور وإنجاب الإناث متساويا الاحتمال ، فإن احتمال ولادة مولود ذكر يكون ح = $\frac{١}{٢}$ وإذا كان هذا الفرض صحيحاً يكون توزيع عدد الذكور في العائلات ذوات الأربعة الأطفال هو توزيع ذى الحدين دليلاه ٤ ، $\frac{١}{٢}$. وكما في المثال (٣ – ٢) يتولد لدينا التوزيع الآتي :

عدد الأطفال الذكور في العائلة س ر	:	٠	١	٢	٣	٤
احتمال هذا العدد ل ر	:	$\frac{١}{١٦}$	$\frac{٤}{١٦}$	$\frac{٦}{١٦}$	$\frac{٤}{١٦}$	$\frac{١}{١٦}$

ونظراً لأن عدد العائلات في العينات المأخوذة ٣٢٠ فإن التكرارات النظرية التي يمكن مقارنتها بالتكرارات المشاهدة تنتج بضرب هذه الاحتمالات في ٣٢٠ ونحصل بذلك على التوزيع التكرارى النظرى الآتي :

عدد الأطفال في العائلة $س_ر$:	٠	١	٢	٣	٤	
العدد المتوقع للعائلات $ق_ر$:	٢٠	٨٠	١٢٠	٨٠	٢٠	(المجموع ٣٢٠)

ونقول هنا أننا قد وفقنا توزيع ذى الحدين للتوزيع التكرارى المعطى .
بمقارنة التكرارات النظرية $ق_ر$ بالتكرارات المشاهدة $ك_ر$ وهى :

$ق_ر$:	٢٠	٨٠	١٢٠	٨٠	٢٠	
$ك_ر$:	١٣	٤٣	٩٢	١١٢	٦٠	(المجموع ٣٢٠)

نجد أن هناك تفاوتاً كبيراً بينها . وعلى فرض توفر شرط العشوائية والاستقلال فإن هذا التفاوت يعزى إلى خطأ الفرض أن ح = $\frac{1}{2}$.

وينبغي أن نشير هنا مرة أخرى إلى أن حكمنا هذا هو حكم ذاتي قد لا يكون هو الحكم الصحيح ، أما الحكم الموضوعى فيستلزم اللجوء إلى أحد الاختبارات الإحصائية المناسبة مثل اختبار χ^2 الذى سندرسه بعد . انظر المثال (٦ — ٨) في البند (٦ — ٧ — ٢) .

تمارين (٣ — ١)

١ — في نوع من أبصال الزهور المعروف أن معدل الإنبات ٩٥٪ . تعبأ هذه الأبصال وتباع في عبوات يحتوى كل منها على ١٠ أبصال . إذا سحب أحد هذه العبوات عشوائياً وزرع ما بها من أبصال فما احتمال كل من الحدثين الآتيين :
(أ) لا تنبت أى بصلة .
(ب) تنبت بصلة واحدة على الأقل ؟

٢ — معدل الإصابة بمرض ما في نوع من البقر ٢٥٪ اختيرت عينة عشوائية من ٨ بقرات . أوجد :

(أ) احتمال أن تكون بقرتان بالضبط مصابتين .

(ب) الوسط الحسابي والتباين لعدد البقرات المصابة في العينات من الحجم ٨ .

لماذا ينبغي أن نفترض هنا أن هذا المرض غير معد للبقر ؟

٣ – احتمال إصابة هدف ثابت ٠,٢ . إذا أطلقت ٥ طلقات مستقلة على هذا الهدف فما احتمال إصابته مرة واحدة على الأقل ؟

٤ – لاختبار الفرض القائل أن نسبة الحصي الجرانيتية إلى مجتمع الحصي على أحد الشواطيء هو ح = ٠,٠٢ أخذت من هذا الشاطيء ١٠٠ عينة عشوائية بكل منها ثلاث حصوات وحسب عدد الحصوات الجرانيتية بكل عينة فنتج التوزيع التكراري المدون في المثال (٣ – ٤) وهو :

عدد الحصوات الجرانيتية في العينة :	٠	١	٢	٣
عدد العينات :	٥٨	٣٣	٧	٢

اختبر ما إذا كانت هذه البيانات تدعم الفرض المذكور على فرض أن توزيع عدد الحصي الجرانيتية هو توزيع ذى الحدين .

(٣ – ٤) توزيع بواسون : POISSON DISTRIBUTION

توزيع بواسون هو توزيع احتمال لمتغير عشوائي وثاب س تأخذ دالة كتلة احتماله الصورة

$$د(س) = \frac{م^{س}}{س!} هـ^{-م} \quad \text{حيث } س = ٠، ١، ٢، ...، م > ٠ \qquad (٨)$$

وحيث هـ أساس اللوغاريتمات الطبيعية (هـ = ٢,٧١٨٢٨ تقريباً)

ولهذه الدالة دليل واحد هو العدد م وبالتالى يتحدد التوزيع تماما إذا عرفت قيمة م . فمثلا حين م = ٢ يكون التوزيع كالآتي :

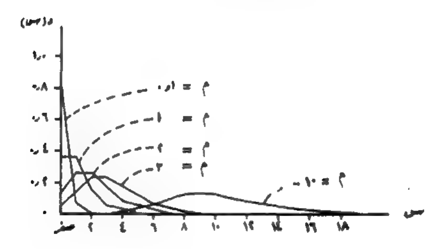
س :	٠	١	٢	٣	٤	٥	٦	..
د (س) :	$هـ^{-٢}$	$٢ هـ^{-٢}$	$٢ هـ^{-٢}$	$\frac{٤}{٣} هـ^{-٢}$	$\frac{٢}{٣} هـ^{-٢}$	$\frac{٤}{١٥} هـ^{-٢}$	$\frac{٤}{٤٥} هـ^{-٢}$	...
أى د(س):	٠,١٣٥	٠,٢٧١	٠,٢٧١	٠,١٨٠	٠,٠٩٠	٠,٠٣٦	٠,٠١٢	...

ومن المميزات الرئيسية لتوزيع بواسون أن

الوسط الحسابي = التباين = م (٩)

ملاحظة (١) :

مجموعة قيم المتغير البواسوني س هي مجموعة لا نهائية {٠ ، ١ ، ٢ ، ...} إلا أنه بعد قيمة معينة س تتوقف على الدليل م ، تتناقص احتمالات هذه القيم تدريجياً حتي تكاد تنعدم كما يشير إلى ذلك الشكل (٣ – ٢) الآتي :



الشكل (٣ – ٢) المضلعات التكرارية لتوزيع بواسون لقيم مختلفة للوسط الحسابي م

ملاحظة (٢) :

هناك جداول تعطى قيم هـ^-م لبعض قيم م – انظر الجدول (٤) بملحق الكتاب – كما أن هناك جداول تعطى الاحتمالات والاحتمالات المتجمعة ل (س = أ) ، ل (س ≤ أ) لتوزيع بواسون لبعض قيم الدليل م والعدد أ – انظر الجدول (٥) بملحق الكتاب .

يستفاد من توزيع بواسون في الموضوعين المقدمين بالبندين الآتيين .

(٣ – ٤ – ١) توزيع بواسون كتقريب لتوزيع ذى الحدين :

نستطيع رياضياً إثبات أنه بوضع م = ح ن في توزيع ذى الحدين المعرف بالدالة (٤) فإنه يؤول إلى توزيع بواسون المعرف بالدالة (٨) حين تقترب ن من اللانهاية وتقترب ح من الصفر .

وهذا يعني من الناحية العملية أنه حين يكون حجم العينة ن كبيراً والاحتمال الثابت ح صغيراً فإن الاحتمالات في توزيع ذى الحدين يمكن إيجادها بالتقريب من الدالة (٨) بدلا من الدالة (٤) مع وضع م = ن ح أى بحيث يكون متوسط توزيع بواسون مساوياً لمتوسط توزيع ذى الحدين . وقد وجد أن هذا التقريب يكون جيداً ، أى يمكن التجاوز عن الخطأ الناشيء عنه إذا كانت :

(ن ⩾ ٥٠ ، ن ح ⩽ ٥) أو (ح ⩽ ٠,١ ، ن ح ⩽ ٥)   (١٠)

إن هذا التقريب من شأنه تبسيط حساب احتمالات ذى الحدين إذ أن حسابها من الدالة (٨) أسهل بكثير من حسابها من الدالة (٤) خاصة إذا كانت ن خارج الحدود التي وضعت لها جداول ذى الحدين .

مثال (٣ – ٦) :

إذا كان احتمال أن يحدث رد فعل سيء للشخص الذى يحقن بمصل ما هو ٠,٠٠١ فاحسب احتمال أن تحدث ٤ حالات ردود فعل سيئة من بين ٢٠٠٠ شخص يحقنون بهذا المصل .

الحل :

توزيع عدد الأشخاص الذين يحدث لهم ردود فعل سيئة هو توزيع ذى الحدين دليلاه ن = ٢٠٠٠ ، ح = ٠,٠٠١ ودالة كتلة احتماله هى :

د (س) = $\text{ق}_{\text{س}}^{٢٠٠٠}$ (٠,٠٠١)$^{\text{س}}$ (٠,٩٩٩)$^{٢٠٠٠-\text{س}}$ حيث س = ٠ ، ١ ، ٢ ، ... ، ٢٠٠٠ .

احتمال ٤ حالات ردود فعل سيئة هو :

$$\text{ل}(\text{س} = ٤) = \text{د}(٤) = \text{ق}^{٢٠٠٠}_{٤}\ (٠,٠٠١)^{٤}\ (٠,٩٩٩)^{١٩٩٦} = ٠,١٠٥٠$$

من الواضح أن حساب هذا الاحتمال لم يكن سهلا فقد تطلب استخدام اللوغاريتمات والحاسب ، غير أنه يمكننا إيجاد الاحتمال المطلوب تقريبياً من توزيع بواسون نظراً لتوفر أحد شروط التقريب وهو أن ن = ٢٠٠٠ أكبر من ٥٠ ، ن ح = ٢ أصغر من خمسة .

$$\text{نضع م} = \text{ن ح} = ٢٠٠٠ \times ٠,٠٠١ = ٢$$

$$\text{د}(\text{س}) = \frac{٢^{\text{س}}}{\text{س}!}\ \text{هـ}^{-٢} \qquad \text{حيث س} = ٠ ، ١ ، ٢ ، \ldots$$

$$= \frac{٢^{٤}}{٤!}\ \text{هـ}^{-٢} = \frac{٢}{٣} \times ٠,١٣٥٣ = ٠,٠٩٠٢$$

وهذا الناتج يمكن إيجاده من الجدول (٥) وهو قريب جداً من الناتج السابق وهو ٠,١٠٥٠ .

(٣ – ٤ – ٢) توزيع بواسون كنموذج لتوزيع الأحداث النادرة :

بصرف النظر عن الدور الذى يلعبه توزيع بواسون كتقريب لتوزيع ذى الحدين فإن له شخصيته واستخداماته الخاصة ، فقد دلت التجربة على أنه يصلح كنموذج احتمال لبعض الأحداث التي تقع عشوائياً عبر الزمان أو المكان .

وعلى سبيل المثال وجد أنه باختيار قيمة مناسبة للدليل م فإن توزيع المتغير س الذى يعبر عن عدد جسيمات ألفا التي تنبعث من مادة مشعة في وحدة زمن مناسبة يمكن أن يعتبر توزيعاً بواسونياً . وبالمثل للمتغيرات التي تعبر (في فترات زمن مناسبة) عن عدد التغيرات الفجائية في الصفات الوراثية – عدد حوادث الطائرات – تتابع طلبات الإغاثة على مراكز الإسعاف – تتابع المكالمات التليفونية على مراكز الهاتف – عدد ولادات ٤ توائم في مدينة – عدد حالات الانفلوانزا

التي ترد إلى مستشفي كبير ... كذلك المتغيرات التي تعبر عن عدد الطحالب في مربع على سفح جبل – عدد الطفيليات على أحد العوائل – عدد البكتريا من نوع معين على طبق بترى Petri plate – عدد الأخطاء المطبعية في صفحة كتاب – الخلل الذى يحدث في جهاز معقد ...

إن مثل هذه الحالات ينظر إليها نمطياً على أنها عملية تولد عدداً من التغيرات أو الأحداث (مثل ظهور جسيم الفا ، طحلب ، بكتريا) في وحدة زمن أو وحدة فراغ مناسبة ، وهذه الوحدة مقسمة إلى عدد كبير جداً من الأجزاء الصغيرة جداً سنسميها **لحظات** instants (سواء كان التقسيم من حيث الزمن أو من حيث الفراغ) وكل من هذه اللحظات يعتبر محاولة يمكن أن يقع فيها الحدث أو لايقع ، وبالتالى فإن هناك إمكانية وقوع الحدث في عدد كبير من المرات .

ويتخذ المتغير توزيعاً بواسونياً إذا توفر الشرطان الآتيان :

(أولا) ندرة الحدث :

يشترط أن يكون معدل وقوع الحدث ، أى متوسط عدد مرات وقوعه في وحدة الزمن أو وحدة الفراغ ، صغيراً بالنسبة لعدد المحاولات التي يمكن أن تسفر عن وقوع الحدث . وهذا ما يجعلنا نصف الحدث بأنه حدث نادر . اعتبر مثلا توزيع المتغير س الذى يعبر عن عدد البكتريا في طبق بترى : إن وحدة الفراغ هنا هى طبق بترى الذى ننظر إليه على أنه مكون من عدد كبير من المساحات الميكروسكوبية (لحظات) كل منها قد يشتمل أو لا يشتمل على بكتريا . وجد بالتجربة أن متوسط عدد البكتريا في الطبق هو عدد متواضع بالرغم من أن هناك بالفعل عدداً لا نهائياً من المحاولات التي يمكن أن تنتج عنها بكتريا ، وعلى هذا فالحدث هو حدث نادر . كذلك اعتبر المتغير الذى يعبر عن عدد الأخطاء في صفحة كتاب جيد الطبع . إن هذه الصفحة تتألف من عدد كبير من الكلمات (لحظات) كل منها قد يقع في كتابتها خطأ أو لا يقع وعلى ذلك فهناك إمكانية

وقوع عدد كبير من الأخطاء ، ولكن نظراً لأن معدل وقوع الخطأ هو عدد صغير جداً نعتبر أن هذا الحدث هو حدث نادر .

واعتبار الحدث نادر هو مسألة نسبية تتطلب أن تكون وحدة الزمن أو وحدة الفراغ كبيرة كبراً كافياً ، فمثلا عندما نعد الطحالب على مربع ما يجب أن يكون هذا المربع كبيراً كبراً كافياً يسمح بنمو عدد وافر من الطحالب مادامت الظروف البيولوجية مهيئة لذلك فلا يجوز مثلا أن تكون مساحة المربع ١ سم^{٢} فقط فهذه المساحة أصغر من جعل الطحالب تتوزع بواسونياً . كذلك في تسجيل حالات الانفلوانزا التي ترد إلى مستشفي كبير لا ينبغى أن تقل وحدة الزمن عن أسبوع مثلا ، لإعطاء الفرصة لورود حالات كافية .

(ثانيا) استقلال الأحداث (عشوائية وقوع الأحداث) :

يشترط أن يكون وقوع الأحداث **عشوائياً** بمعنى أن يكون احتمال وقوع أو عدم وقوع الحدث في أى لحظة مستقلا عن وقوعه أو عدم وقوعه في أى لحظة سابقة أو لاحقه غير متداخلة معها وبذلك لا يتأثر وقوع الحدث إلا بالعوامل العشوائية وحدها . فمثلا وجود طحلب في جزء من مربع ما لا ينبغى أن يزيد أو ينقص من احتمال نمو طحالب أخرى في أى جزء آخر من المربع . كذلك تسجيل حالة انفلوانزا في لحظة ما لا يجب أن يؤثر في احتمال تسجيل حالات تالية .

تحت هذين الشرطين اتضح أنه بتقريب جيد إلى حد كبير أو بالضبط يمكن إيجاد احتمال عدد مرات وقوع الحدث في أى فترة زمنية أو فراغية بواسطة توزيع بواسون دليله :

$$م = ك\ ن \qquad (11)$$

أى من الدالة $د(س) = \frac{(ك\ ن)^{س}}{س!} هـ^{-ك\ ن}$ حيث $س = ٠ ، ١ ، ٢ ، ....$

وحيث د (س) هو احتمال وقوع الحدث س من المرات خلال فترة زمنية أو فراغية ز وحيث ك مقدار ثابت موجب يعبر عن متوسط وقوع الحدث في وحدة الزمن أو الفراغ ، وهذا المقدار يحسب تجريبياً .

مثال (٣ – ٧) :

إذا فرض أن البكتريا من نوع معين تتواجد في الماء بمعدل ٢ بكتريا في السنتيمتر المكعب وأن عدد البكتريا يتوزع توزيعاً بواسونيا فأوجد احتمال أنه في عينة عشوائية من ٢ سم^{٣} من الماء : (أ) لا يوجد بكتريا (ب) يوجد ٣ بكتريا على الأقل .

الحل :

لدينا ك = ٢ بكتريا في السنتيمتر المكعب ، ز = ٢ سم^{٣} .

إذن م = ٢ × ٢ = ٤

، د (س) = $\frac{٤}{س!}$ هـ$^{-٤}$ حيث س = ٠ ، ١ ، ٢ ، ...

(أ) احتمال عدم وجود بكتريا في ٢ سم^{٣} = د (٠) = هـ$^{-٤}$ = ٠,٠١٨٣٢

(ب) احتمال وجود ٣ بكتريا على الأقل في ٢ سم^{٣} = ل (س ≥ ٣) .

= ١ – ل (س ≤ ٢) = ١ – [د (٠) + د (١) + د (٢)]

= ١ – (١ + ٤ + ٨) هـ$^{-٤}$ = ١ – ١٣ هـ$^{-٤}$.

= ٠,٧٦١٨٤

مثال (٣ – ٨) :

إذا كان متوسط عدد السيكلوب في لتر من ماء بحيرة هو ٢ فما احتمال وجود ٥ سيكلوب على الأقل في عينة من ٣ لترات من ماء البحيرة ؟

(السيكلوب كائن دقيق يطفو على الماء وتقتات عليه الأسماك) .

الحل :

لدينا ك = ٢ سيكلوب في اللتر ، ز = ٣ لتر

إذن م = ٢ × ٣ = ٦

، د (س) = $\frac{٦^{س}}{س!}$ هـ$^{-٦}$ حيث س = ٠ ، ١ ، ٢ ، ....

احتمال وجود ٥ سيكلوب على الأقل = ل (س ≥ ٥) = ١ − ل (س ≤ ٤)

= ١ − ٠,٢٨٥

= ٠,٧١٥

(٣ – ٤ – ٣) اختبار استقلال الأحداث النادرة (أو اختبار العشوائية) :

عند تناول حدث نادر يهمنا في كثير من الأحيان دراسة استقلال الأحداث أى دراسة ما إذا كان وقوع الحدث في لحظة ما يزيد أو ينقص من احتمال وقوعه في لحظة تالية كما هو الحال مثلا في دراسة توزيع سوسة الفاصوليا أو توزيع حشرة على نوع من الذباب . ونظراً لأن الحدث النادر لا يتوزع بواسونيا إلا إذا كانت الأحداث مستقلة فإننا نستطيع الحكم على استقلال الأحداث عن طريق اختبار ما إذا كان التوزيع بواسونيا ، وهذا الاختبار يمكن إجراؤه بتوفيق توزيع بواسون لتوزيع تكرارى مشاهد في تجربة كما فعلنا في حالة ذى الحدين في البند (٢ – ٣ – ١) والمثال (٣ – ٤) . فإذا كانت المطابقة حسنة دل ذلك على أن التوزيع بواسونيا وبالتالى تكون الأحداث مستقلة وتقع عشوائياً ، أما إذا لم تكن المطابقة حسنة فنحكم بعدم عشوائية وقوع الأحداث .

مثال (٣ – ٩) :

أجريت تجربة لاختبار توزيع خلايا الخميرة في ٤٠٠ مربع من جهاز الـ hemacytometer (صندوق لعد الخلايا) ووجد التوزيع التكرارى المبين بالعمودين الأول والثانى من الجدول (٣ – ٢) . بالتأمل في هذين العمودين نلاحظ أمرين هامين هما :

(أ) أن ٧٥ من هذه المربعات أى حوالى ١٩٪ منها لا تشتمل على أى خلية ومعظم المربعات (٥٦٪) تحمل إما خلية واحدة أو خليتين وأن ١٧ مربعاً فقط (أى حوالى ٤٪) تحتوى على ٥ خلايا أو أكثر .

(ب) الوسط الحسابي $\bar{س} = \frac{١}{ن} مج \, ك_ر \, س_ر$

$$= \frac{١}{٤٠٠} (٧٥ \times ٠ + ١٠٣ \times ١ + ١٢١ \times ٢ \times ... + ١ \times ٩)$$

$$= \frac{٧٢٠}{٤٠٠} = ١,٨$$

الجدول (٣ − ٢)

التكرارات المشاهدة والتكرارات البواسونية المتوقعة لعدد خلايا الخميرة في ٤٠٠ مربع من جهاز الهيماسيتومتر

عدد الخلايا في المربع س	التكرارات المشاهدة ك	التكرارات النسبية المتوقعة ل = د (س)	التكرارات المتوقعة و=ل×٤٠٠	الانحرافات ك − و
٠	٧٥	٠,١٦٥٣	٦٦,١	+
١	١٠٣	٠,٢٩٧٥	١١٩,٠	−
٢	١٢١	٠,٢٦٧٨	١٠٧,١	+
٣	٥٤	٠,١٦٠٧	٦٤,٣	−
٤	٣٠	٠,٠٧٢٣	٢٨,٩	+
٥	١٣ } ١٧	٠,٠٢٦٠	١٠,٤ } ١٤,٥	+ } +
٦	٢	٠,٠٠٧٨	٣,١	−
٧	١	٠,٠٠٢٠	٠,٨	+
٨	٠	٠,٠٠٠٤	٠,٢	−
٩	١	٠,٠٠٠١	٠,٠	−
	٤٠٠	٠,٩٩٩٩	٣٩٩,٩	

أى أن متوسط عدد الخلايا في المربع هو ١,٨ خلية وهذا عدد صغير فعلا بالنسبة لسعة كل مربع وبالنسبة لعدد الخلايا التي يحتمل أن تظهر في أى من المربعات .

من هذا يحق لنا أن نعتبر أن الحدث هو حدث نادر ونتوقع بالتالى أن يكون توزيعه بواسونيا إذا توفر شرط الاستقلال . ولاختبار هذا الشرط نوفق توزيع بواسون للتوزيع التكرارى المشاهد مع تقدير الدليل م لذلك التوزيع من العينة أى نأخذ م = ١,٨ فتكون دالة كتلة الاحتمال :

$$\text{د}\,(\text{س}) = \frac{\text{١,٨}^{\text{س}}}{\text{س}!}\,\text{هـ}^{-\text{١,٨}} \quad \text{حيث س} = \text{٠ ، ١ ، ٢ ، ....}$$

نحسب الاحتمالات د (٠) ، د (١) ، د (٢) ، ... كما في العمود الثالث من الجدول (٣ – ٢) ثم نضرب كلا من هذه الاحتمالات (التكرارات النسبية المتوقعة) في حجم التوزيع التكرارى المشاهد وهو ٤٠٠ فنحصل على التكرارات المتوقعة المبينة بالعمود الرابع

بمقارنة التكرارات المشاهدة بالتكرارات النظرية نجد أن هناك مطابقة حسنة بين التوزيع التكرارى المشاهد وتوزيع بواسون دليله ١,٨ ولا يوجد نمط واضح لانحرافات التكرارات المشاهدة عن التكرارات المتوقعة لها كما يبدو من العمود الخامس وإن كان الحكم الموضوعي لهذا التطابق لا يتأتى إلا بأحد الاختبارات الإحصائية التي سندرسها بعد . انظر المسألة (٦) في تمارين (٦ – ٢)

ونستنتج من هذا أن توزيع خلايا الخميرة هو توزيع بواسوني وهذا تتضمن أن الاحداث هنا تقع عشوائيا أى مستقلة عن بعضها .

وهناك اختبار آخر يساعد على بيان ما إذا كان التوزيع بواسونيا دون الالتجاء إلى عملية التوفيق . ويعتمد هذا الاختبار على الخاصة الهامة التي جاءت في المتساوية (٩) عن تساوى التباين والوسط الحسابي في التوزيعات البواسونية . وحين نتناول

عينة عشوائية من مجتمع بواسوني نتوقع وجود هذا التساوى بالتقريب أى نتوقع أن تكون النسبة بين التباين والوسط الحسابي المحسوبين من العينة قريبة من الواحد الصحيح . في المثال الأخير نجد أن :

$$\text{تباين العينة } ع^{٢} = \frac{١}{ن - ١}\left[\text{مج } ك_{ر}\, س_{ر}^{٢} - \frac{(\text{مج } ك_{ر}\, س_{ر})^{٢}}{ن}\right]$$

$$= \frac{١}{٣٩٩}\left[(٨١\times١+ \ldots +٤\times١٢١+١\times١٠٣+٠\times٧٥) - \frac{٧٢٠^{٢}}{٤٠٠}\right]$$

$$= ١,٩٦٥$$

وهذا العدد قريب من الوسط الحسابي ١,٨ كما أن النسبة بينهما $\frac{١,٩٦٥}{١,٨} = ١,٠٩٢$ قريبة من الواحد وهذا يدعم استنتاجنا السابق بأن توزيع المتغير هو توزيع بواسوني وما يتبع ذلك من عشوائية الأحداث . هذا ويمكن اختبار كبر أو صغر النسبة $ع^{٢} / \bar{س}$ عن الواحد بطريقة موضوعية عن طريق اختبار ت الذى سندرسه بعد . انظر المثال (٦ – ٥) بالبند (٦ – ٦ – ٤) .

نمط التجمع ونمط التنافر :

إذا اتضح أن التكرارات المشاهدة تنحرف عن التكرارات المتوقعة لها بشكل جوهرى أو أن النسبة $ع^{٢} / \bar{س}$ أكبر أو أصغر من الواحد بشكل جوهرى فإن هذا يعني أن الأحداث لا تقع مستقلة عن بعضها بل يؤثر وقوع أو عدم وقوع أحدها في وقوع أو عدم وقوع الأحداث الأخرى . وهذا يدعو إلى التساؤل عما إذا كانت التكرارات المشاهدة تنم عن نمط خاص وعن تفسير ما قد يوجد من أنماط . وهذا التفسير لا يستطيع التحليل الإحصائي وحده القيام به فالمرجع الأول في هذا هو معرفة الباحث بظروف التجربة وطبيعة المتغيرات والعوامل التي تؤثر فيها .

وهناك نوعان رئيسيان من الأنماط هما :

(١) نمط التجمع : **CLUMPING**

في هذا النمط تكون التكرارات المشاهدة أكبر من التكرارات المتوقعة عند ذيلى التوزيع وأصغر منها عند الوسط كما هو الحال في التجربة الملخصة بالجدول (٣ – ٣) الآتي الذى يعرض توزيع الحلم المائي water mite على ٥٨٩ هاموشة . ويوصف هذا النمط بأنه مُعْدى contagious بمعني أن وقوع حدث (ظهور حلمة مثلا) يرفع من احتمال وقوع أحداث أخرى والعكس بالعكس . وفي هذه الحالة تكون النسبة بين التباين والوسط الحسابي في المجتمع أكبر من الواحد .

(٢) نمط التنافر : **REPULSION**

في هذا النمط تكون التكرارات المشاهدة أصغر من التكرارات المتوقعة عند الذيلين وأكبر منها عند الوسط كما هو الحال في التجربة الملخصة بالجدول (٣ – ٤) الذى يسجل توزيع السوس على ١١٢ نبات فاصوليا . وهنا يكون وقوع الحدث (خروج سوسة مثلا) عائقاً لوقوع أحداث أخرى فيشتمل التوزيع على عدد قليل من المجموعات المتجانسة وعدد كبير من المجموعات المختلطة . وفي هذه الحالة تكون النسبة بين التباين والوسط الحسابي في المجتمع أصغر من الواحد .

بعد حساب الوسط الحسابي والتباين للتوزيع المشاهد نجد ما يلى :

بالنسبة للتجربة الأولى : ع^{٢} / $\overline{س}$ = ٠,٩٧٠٩ / ٠,٤٣٤٣ = ٢,٢٣٦

هذه النسبة أكبر جوهرياً من الواحد . انظر البند (٦ – ٦ – ٤) .

بالنسبة للتجربة الثانية : ع^{٢} / $\overline{س}$ = ٠,٢٦٩ / ٠,٤٦٤٣ = ٠,٥٧٩

هذه النسبة أصغر جوهريا من الواحد . انظر البند (٦ – ٦ – ٤) .

الجدول (٣-٣)

التكرارات المشاهدة والتكرارات البواسونية المتوقعة

لعدد الحلم على ٥٨٩ هاموشه (نمط تجمع)

عدد الحلم على الهاموشة	ك	ق	ك − ق
٠	٤٤٢	٣٨٠,٧	+
١	٩١	١٦٦,١	−
٢	٢٩	٣٦,٢	−
٣	١٤	٥,٣	+
٤	٤	٠,٦	+
٥	٦ (٢٧ } ٣–٨)	٠,١ (٦ } ٣–٨)	+ (+ } ٣–٨)
٦	٢	٠,٠	−
٧	٠	٠,٠	٠
٨	١	٠,	|
	٥٨٩	٥٨٩	

جدول (٣-٤)

التكرارات المشاهدة والتكرارات البواسونية المتوقعة

لعدد السوس الذى خرج من ١١٢ نبات فاصوليا (نمط تنافر)

عدد السوس على النبات	ك	ق	ك − ق
٠	٦١	٧٠,٤	−
١	٥٠	٣٢,٧	+
٢	١	٧,٦	−
٣	(١ } ٢–٤)	١,٢ (٨,٩ } ٢–٤)	− (− } ٢–٤)
٤	٩	٠,١	−
	١١٢	١١٢	

تمارين (٣ - ٢)

(١) في توزيع بواسون دليله م = ٠,٧٢ أوجد :

ل (س = ٠) ، ل (س = ١) ، ل (س = ٢) ، ل (س > ٢) .

(اعتبر أن هـ$^{-٧٢}$= ٠,٤٨٦٨)

(٢) دلت الخبرة الطويلة على أن السفن تدخل في إحدى الموانى بمعدل ٣ سفن في الساعة . إذا كان توزيع عدد السفن التي تدخل هذا الميناء هو توزيع بواسون فأوجد احتمال أنه في فترة زمنية قدرها دقيقتان (أ) لا تدخل أى سفينة (ب) تدخل سفينتان على الأقل .

(٣) إذا علم أن الجسيمات تنبعث من مصدر مشع بمعدل ٠,٥ جسيماً في الثانية وأن عدد هذه الجسيمات يتوزع بواسونياً فاحسب احتمال انبعاث ٣ جسيمات أو أكثر في فترة زمنية طولها ٦ ثواني .

(٤) الجدول الآتي يعرض توزيع عدد الحلم (المايت المائية) water mite على نوع من الذباب . وفق توزيعاً بواسونياً لهذا التوزيع واستنتج أن وقوع الأحداث (ظهور الحشرات على الذبابة) ليس عشوائياً . أوجد تباين التوزيع التكرارى المشاهد وقارنه بالوسط الحسابي لتدعيم استنتاجك (ستجد أن النسبة بين التباين والوسط الحسابي ٢,٢٢٥) .

عدد الحشرات على الذبابة	٠	١	٢	٣	٤	٥	٦	٧	٨	المجموع
التكرارات المشاهدة	٤٢٢	٩١	٢٩	١٤	٤	٦	٢	٠	١	٥٨٩

(٥) أجب عن نفس السؤال السابق مستخدماً التوزيع التكرارى الآتي الذى يعرض توزيع عدد السوس على قرن فاصوليا (قيس هذا العدد بعدد الثقوب التي نتجت في قرن الفاصوليا أثناء خروج اليرقات منها) :

عدد السوس في القرن :	٠	١	٢	٣	٤	المجموع
التكرارات المشاهدة :	٦١	٥٠	١	٠	٠	١١٢

(٦) يدخل الزبائن في أحد المحلات بمعدل ٣٠ شخصاً في الساعة فإذا كان توزيع عدد الزبائن هو توزيع بواسوني فأوجد احتمال أنه في فترة زمنية قدرها دقيقتان (أ) لا يدخل أحد (ب) يدخل شخصان على الأقل .

(٣ - ٥) توزيع باسكال : PASCAL DISTRIBUTION

في توزيع ذى الحدين نعتبر أن حجم العينة هو عدد ثابت ن ونعالج متغيراً عشوائياً س يعبر عن عدد مرات وقوع الحدث في ن من المحاولات ، إلا أن هناك مشكلات تتطلب عكس هذا الوضع فيكون حجم العينة متغيراً عشوائياً س ويكون عدد مرات وقوع الحدث هو عدد ثابت أ محدد من قبل ويكون المطلوب هو إيجاد احتمالات قيم المتغير س التي تسمح بوقوع الحدث هذا العدد المحدد من المرات . (يلاحظ أن المعاينة هنا تكون من النوع التتابعى sequential sampling)

أى أننا نعالج متغيراً عشوائياً وثابتاً س يعبر عن العدد اللازم من المحاولات لكى يقع الحدث عدداً محددا أ من المرات .

وبطبيعة الحال لا يجب أن يقل عدد المحاولات عن العدد أ لأن وقوع الحدث أ من المرات يتطلب أ محاولة على الأقل ، وعلى ذلك تبدأ قيم س بالعدد أ . أى أن هذا المتغير لا يأخذ إلا القيم أ ، أ + ١ ، أ + ٢ ، ... فإذا كان المطلوب وقوع الحدث ٤ مرات مثلا فإن س تأخذ القيم ٤ ، ٥ ، ٦ ، ...

تحت نفس افتراضات العشوائية وثبات الدليل ح واستقلال الأحداث يمكن أن نثبت رياضياً أن دالة كتلة الاحتمال لهذا المتغير تأخذ الصورة الآتية :

$$د(س) = {}^{س-١}ق_{أ-١}\; ح^{أ}\; ك^{س-أ} \qquad \text{حيث } س = أ ، أ+١ ، أ+٢ ، ... \qquad (١٢)$$

وحيث ١ > ح > ٠ ، ك = ١ − ح ، أ مقدار ثابت

ويسمى توزيع الاحتمال حينئذ بتوزيع باسكال دليلاه ح ، أ كما يسمى المتغير س بمتغير باسكال . يلاحظ أن الحدث يقع أ من المرات المستقله باحتمال ح فى كل مرة وأنه يفشل فى س – أ من المرات المستقلة باحتمال ك فى كل مرة ، وأن المرة الأخيرة هى نجاح دائما . كما يلاحظ أن عدد الطرق التى تقع فيها (أ – ١) نجاحات من (س – ١) محاولات هو $ق^{س-١}_{أ-١}$ وعلى ذلك فإن احتمال وقوع الحدث أ من المرات (وفشله س – أ من المرات) هو $ق^{س-١}_{أ-١}\ ح^{أ}\ ك^{س-أ}$

مثال (٣ – ١٠) :

المعروف أن ٦٠٪ من المرضي بمرض معين يستجيبون للدواء ما بعد تناوله لمدة أسبوع . يختار مرضي بهذا المرض الواحد بعد الآخر في ترتيب عشوائي ليتناولوا الدواء (لمدة أسبوع) حتي نحصل على ٥ استجابات صحيحة :

(أ) ما احتمال أن يكون العدد المختار من المرضي سبعة ؟

(ب) ما احتمال أن يكون العدد المختار من المرضي عشرة ؟

الحل :

إن عدد المرضي اللازمين للحصول على خمس استجابات صحيحة هو متغير عشوائي س له توزيع باسكال دليلاه ٠,٦ ، ٥ ودالة كتلة احتماله :

$$د(س) = ق^{س-١}_{٤}\ (٠,٦)^{٥}\ (٠,٤)^{س-٥} \qquad \text{حيث س = ٥ ، ٦ ، ٧ ، ....}$$

$$(أ)\ ل\ (س=٧) \qquad = د(٧) = ق^{٦}_{٤}\ (٠,٦)^{٥}\ (٠,٤)^{٢}$$

$$= ٠,١٨٦٦$$

$$(ب)\ ل\ (س=١٠) = د(١٠) = ق^{٩}_{٤}\ (٠,٦)^{٥}\ (٠,٤)^{٥}$$

$$= ٠,١٠٠٣$$

ملاحظة :

إن الحالات التي يظهر فيها متغير باسكالى تنشأ فى المعتاد عندما يستخدم ما يسمى بالمعاينة التتابعية sequential sampling حيث لا يحدد حجم العينة مسبقا ، بل تختار المشاهدات فى تتابع عشوائى الواحدة بعد الأخرى وتتوقف هذه العملية حين يتجمع عدد كاف من المشاهدات يمكننا من اتخاذ قرار بحسب قاعدة معينة توضع سلفا . ففى المثال (٣ — ١٠) نفرض أن المطلوب اختبار صحة الفرض أن ٦٠٪ من المرضى يستجيبون للدواء ، وأن القاعدة التى وضعت لتحديد صحة أو خطأ هذا الفرض كالآتى :

« اختر المرضى الواحد بعد الآخر بترتيب عشوائى ليتناول الدواء وسجل العدد سَ لعدد المرضى المختبرين حتى تحصل على ٥ استجابات صحيحة . ارفض الفرض إذا كان الاحتمال ل (س ≥ سَ) يساوى أو يقل عن ٠,٠٥ وإلا فاقبل الفرض » .

إذا استخدمت هذه القاعدة فماذا يكون حكمنا عن الفرض إذا وجد فى تجربة ما أن عدد المرضى المختبرين حتى الوصول إلى ٥ استجابات صحيحة هو :

(أولا) سَ = ٨ (ثانيا) سَ = ١٣

الحل :

(أولاً) ل(س ≥٨) = ١− ل(س <٨) = ١− [د(٥)+ د(٦)+ د(٧)]

=١−[(٠,٦)^٥+٥×(٠,٦)^٥×٠,٤+١٥×(٠,٦)^٥×(٠,٤)^٢]

= ١−(٠,٦)^٥ [١+٢+٢,٤] = ٠,٥٨

بما أن ٠,٥٨ أكبر من ٠,٠٥ نقبل الفرض أن نسبة الشفاء ٦٠٪

(ثانياً) ل(س≥١٣)=١−ل(س<١٣)=١−[د(٥)+د(٦)+...+د(١٢)]

==١−(٠,٦)^٥[١+٢+٢,٤+...+٠,٣٢٤]=٠,٠٣

بما أن ٠,٠٣ أصغر من ٠,٠٥ نرفض الفرض أن نسبة الشفاء ٦٠٪ .

(٣ – ٦) التوزيع الهندسي : THE GEOMETRIC DISTRIBUTION

هو حالة خاصة من توزيع باسكال تكون فيها عدد مرات وقوع الحدث أ = ١ ويرمز المتغير سـ هنا إلى عدد المحاولات اللازمة لوقوع الحدث لأول مرة . وتنتج دالة كتلة الاحتمال لهذا المتغير بوضع أ = ١ في الصيغة (١٢) أى تكون على الصورة :

$$د (س) = ح ك^{س-١} \qquad حيث\ س = ١ ، ٢ ، ٣ ، ... \qquad (١٣)$$

ويسمى التوزيع حينئذ بالتوزيع الهندسي أو بتوزيع وقت الانتظار waiting time distribution ولهذا التوزيع دليل واحد هو ح . وهو يفيد في دراسة الخواص النادرة للمجتمعات كما هو الحال في أمراض الدم النادرة حيث قد لا ينفعنا استخدام توزيع ذى الحدين ، لأننا لو حددنا حجم العينة فقد لا يقع الحدث في أى عنصر منها وبذلك لا نحصل على معلومات كافية عن الحدث ، بينما التوزيع الهندسي يضمن وقوع الحدث .

يمكن أن نثبت رياضياً أن :

$$الوسط\ الحسابي\ للتوزيع = \frac{١}{ح} \qquad (١٤)$$

$$،\ تباين\ التوزيع = \frac{ك}{ح^{٢}} \qquad (١٥)$$

مثال (٣ – ١١) :

في إحدى المنتجات الصناعية المعروف أنه في المتوسط توجد وحدة معيبة في كل ١٠٠ وحدة . تختار وحدات عشوائياً وتختبر الواحدة بعد الأخرى إلى أن تظهر أول وحدة معيبة . (أ) ما احتمال اختبار ٥ وحدات حتي الوصول إلى الوحدة المعيبة ؟ (ب) ما العدد المتوقع للاختبارات اللازمة للعثور على أول وحدة معيبة ؟

الحل :

لدينا توزيع هندسي دليله ح = ٠,٠١ وإذن دالة كتلة احتماله

د (س) = ح ك$^{س-١}$ = ٠,٠١ (٠,٩٩)$^{س-١}$ حيث س = ١ ، ٢ ، ٣ ، ...

(أ) د (٥) = ٠,٠١ (٠,٩٩)^{٤}

= ٠,٠٠٩٦

(ب) العدد المتوقع يعني الوسط الحسابي = $\frac{١}{ح}$ = $\frac{١}{٠,٠١}$ = ١٠٠ اختبار .

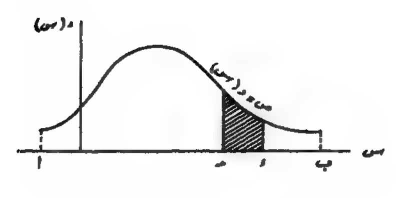
(٣ – ٧) توزيعات الاحتمال المتصلة :

تمتد الأفكار السابقة عن توزيعات الاحتمال للمتغيرات الوثابة إلى حالة المتغيرات المتصلة مع بعض الفروق التي تقتضيها طبيعة كل من هذين النوعين من المتغيرات . فإذا كان سـ متغيراً حقيقياً من النوع المتصل مداه الفترة (أ ، ب) فإن احتمال أن يأخذ هذا المتغير قيمة محددة س يساوى صفرا ، ذلك لأن أى متغير متصل يأخذ عدداً لا نهائياً من القيم في أى جزء من مداه مهما كان صغيراً .

ولذلك يعرف توزيع الاحتمال في هذه الحالة بواسطة دالة متصلة غير سالبة د حيث :

$$ل(س < سـ \leq س + \triangle س) = \int_{س}^{س+\triangle س} د(س) . ءس \qquad (١٦)$$

وهذه الصيغة تعني أن احتمال وقوع قيم المتغير سـ في فترة △ س يساوى تكامل الدالة د على هذه الفترة . إن مثل هذه الدالة تسمى **بدالة كثافة الاحتمال** probability density function وتمثل بيانياً بالشكل (٣ – ٣) .



الشكل (٣ – ٣) توزيع الاحتمال لمتغير متصل

ومن الخواص الرئيسية للمنحني الممثل لأى دالة كثافة احتمال أن مساحة المنطقة الواقعة أسفله وفوق المحور السيني تساوى الواحد الصحيح ، ويمكن رؤية هذا المنحني كخط ممهد لمضلع تكرارى يمثل التكرارات النسبية لتوزيع تكرارى ذى فئات مبني على عدد كبير جداً من المشاهدات موزعة على عدد كبير جداً من الفئات ذوات الأطوال الصغيرة جداً .

ومن تعريف الدالة د نرى أن احتمال وقوع قيم المتغير س بين عددين جـ ، د أى في فترة (جـ ، د) محتواة في المدى (أ ، ب) يعطى بالتكامل .

$$ل(جـ < س \leq د) = \int_{جـ}^{د} د(س) \,.\, ءس \qquad (١٧)$$

أى أننا إذا اخترنا عشوائياً قيمة واحدة من قيم المتغير فإن احتمال وقوعها بين عددين جـ ، د يعطى بالتكامل المذكور . ومن الواضح أن هذا التكامل يساوى عددياً مساحة الجزء المظلل بالشكل (٣ ـــ ٣) ، كما أنه يعبر عن نسبة قيم المتغير الواقعة بين العددين جـ ، د .

ويعرف الوسط الحسابي μ والتباين σ^2 لتوزيع احتمال متغير متصل مداه الفترة (أ ، ب) كالآتي :

$$\mu = \int_{\text{أ}}^{\text{ب}} \text{س}\ \text{د}(\text{س}) .\ \text{ء س} \qquad (18)$$

$$\text{،}\ \sigma^2 = \int_{\text{أ}}^{\text{ب}} (\text{س} - \mu)^2\ \text{د}(\text{س}) .\ \text{ء س} \qquad (19)$$

$$\text{أو} \qquad \int_{\text{أ}}^{\text{ب}} \text{س}^2\ \text{د}(\text{س}) .\ \text{ء س} - \mu^2$$

مثال (٣ – ١٢) :

أوجد الوسط الحسابي والتباين لتوزيع احتمال المتغير المتصل الذى دالة كثافة احتماله هى :

$$\text{د}(\text{س}) = 6\ \text{س}\ (1 - \text{س}) \qquad \text{حيث}\ 1 > \text{س} > 0 .$$

الحل :

$$\mu = \int_{0}^{1} 6\ \text{س}^2\ (1 - \text{س})\ \text{ء س} = \frac{1}{2}$$

$$\sigma^2 = \int_{0}^{1} 6\ \text{س}^3\ (1 - \text{س})\ \text{ء س} - \frac{1}{4} = \frac{1}{20}$$

ومن أشهر توزيعات الاحتمال المتصلة وأكثرها استخداماً ذلك التوزيع المسمى بالتوزيع المعتدل الذى يلعب دوراً رئيسياً في النظرية الإحصائية ونظرية الاحتمال كما أن هناك توزيعات احتمالات متصلة أخرى تستخدم كنماذج احتمال لبعض المجتمعات ، منها التوزيع المعتدل اللوغاريتمى والتوزيع المعتدل الدائرى وتوزيع جاما والتوزيع الأسى ... وسوف نكتفى بتقديم التوزيع المعتدل والتوزيع المعتدل اللوغاريتمى فى الفصل القادم .

الفصل الرابع

التوزيع المعتدل والتوزيع المعتدل اللوغاريتمى

THE NORMAL DISTRIBUTION
AND
THE LOGNORMAL DISTRIBUTION

(أولا) التوزيع المعتدل

التوزيع المعتدل هو أهم توزيعات الاحتمال القياسية وأكثرها استخداما ، إذ يؤخذ كنموذج لكثير من المتغيرات المتصلة ، منها أطوال بتلات بعض النباتات – أطوال أجنحة الذباب المنزلى – أطوال وموازين الأطفال عند الولادة – مقادير الدسم فى الزبد الناتج من بعض أنواع الأبقار ... فضلا عن أنه يلعب دورا كبيرا فى بناء نظرية الاحتمال والنظرية الإحصائية .

وقد سمى هذا التوزيع بالتوزيع المعتدل (أو المعتاد أو الطبيعى) لأنه كان يظن فيما مضى أن أية بيانات عن ظواهر الحياة ينبغى أن تمتثل لهذا التوزيع وإلا كانت هذه البيانات مشكوكا فيها . إلا أنه ثبت الآن أن الأمر ليس كذلك فهناك كثير من الظواهر ذات توزيعات تختلف عن التوزيع المعتدل . كما يسمى التوزيع بتوزيع جاوس اعترافا بفضل العالم الألمانى كارل فردريك جاوس (١٧٧٧ – ١٨٥٥) الذى استنبط التوزيع رياضيا كتوزيع احتمال أخطاء القياس وكان لذلك يسميه

بالقانون الطبيعى للأخطاء normal law of errors ، ويسمى أيضا بتوزيع جاوس – لابلاس اعترافا بجميل العالم الفرنسى بيير سيمون لابلاس (١٧٤٩ – ١٨٢٧) الذى كان أول من اكتشفه وأثبت صلاحيته كنموذج لكثير من الظواهر .

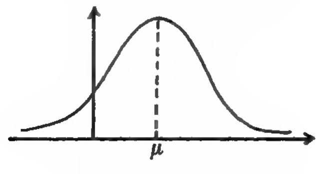
ويعرّف هذا التوزيع بواسطة دالة كثافة الاحتمال المعطاة بالقاعدة :

$$\text{د}(\text{س}) = \frac{1}{\sqrt{2\text{ط}\,\sigma^{2}}}\ \text{هـ}^{-\frac{1}{2\sigma^{2}}(\text{س}-\mu)^{2}} \quad \text{حيث} \quad \infty > \text{س} > -\infty ,\ \sigma > 0 \qquad (1)$$

وهذه الدالة بارامتران هما μ ، σ ، تعبر μ عن الوسط الحسابى وتعبر σ عن الانحراف المعيارى للتوزيع . ويتحدد التوزيع تماما إذا علمت قيمتا هذين الدليلين ، ولذا نرمز له بالرمز مع (μ ، σ) . والمتغير العشوائى س الذى له هذه الدالة يسمى بالمتغير المعتدل .

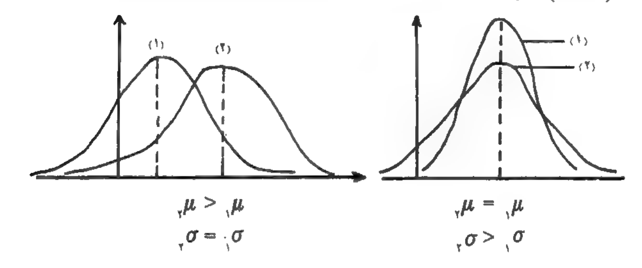
(٤ – ١) بعض خواص التوزيع

(أ) المنحنى الذى يمثل الدالة (١) يسمى بالمنحنى المعتدل وهو منحنى ذو قمة واحدة ومتماثل حول الخط س = μ . ومساحة المنطقة الواقعة بين هذا المنحنى ومحور السينات تساوى الواحد الصحيح كما هو الحال لمنحنى أى توزيع احتمال . ومن التماثل نجد أن الخط س = μ يقسم الشكل إلى منطقتين متساويتى المساحة ومساحة كل منهما تساوى $\frac{1}{2}$. انظر الشكل (٤ – ١) .



الشكل (٤ - ١) : منحنى التوزيع المعتدل

(ب) لكل من البارامترين μ ، σ عدد غير منتهى من القيم ، ولذلك هناك عدد غير منتهى من التوزيعات المعتدلة . هذا مع ملاحظة أن μ هو بارامتر موضع location parameter أما σ فهو بارامتر شكل shape parameter كما يتبين من الشكل (٤ — ٢) الآتى :



الشكل (٤ — ٢)

(جـ) معامل الالتواء = صفر ، معامل التفرطح = ٣

(د) إذا كانت س ترمز إلى قيم متغير معتدل س وسطه الحسابي μ وانحرافه المعيارى σ فإن الصيغة :

$$ع = \frac{س - \mu}{\sigma} \qquad (٢)$$

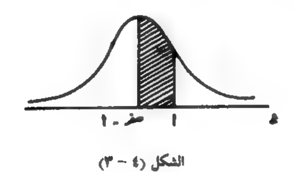
تحول المتغير س إلى متغير ع له توزيع معتدل وسطه الحسابي صفر وانحرافه المعيارى الواحد الصحيح ، ويسمى هذا المتغير بالمتغير المعتدل المعيارى وسنرمز له بالرمز : مع (٠ ، ١) .

(هـ) إذا رسم توزيع التكرارات النسبية المتجمعة لمتغير معتدل على ورق الرسم البياني ذى التقسيم الخطى (العادى) فإن هذا المنحني يتخذ الشكل S غير أن

هناك ورق يسمى ورق تقسيم الاحتمالات المعتدلة **normal probability graph paper** إذا رسم عليه التوزيع نتج خط مستقيم ، ويستخدم هذا الورق للكشف عن اعتدالية المجتمعات كما سنرى في البند (٤ – ٤) الآتي .

(٤ – ٢) جداول المساحات أسفل المنحني المعتدل المعيارى :

نظراً لأهمية معرفة احتمالات المتغيرات المعتدلة وكثرة الحاجة إليها فقد حسبت قيم هذه الاحتمالات لمختلف فترات المتغير المعتدل المعيارى ع ووضعت في جداول تعرف بجداول المساحات أسفل المنحني المعتدل المعيارى ، وهذه الجداول تأخذ صوراً متعددة تؤدى إلى نفس النتائج ومنها الصورة الواردة بالجدول (٦) بملحق هذا الكتاب . وهذا الجدول يعطى المساحة تحت المنحني المعتدل المعيارى وفوق الفترة بين الصفر وبعض أعداد موجبة أ متوقعة للمتغير ع ، وهذه هى المساحة المظللة بالشكل (٤ – ٣) وهى تعبر عن الاحتمال ل (أ ≥ ع > ٠)   (٣)



الشكل (٤ – ٣)

أى عن احتمال وقوع قيم المتغير ع بين العددين . ، أ أى في الفترة (. ، أ) . ويلاحظ من تماثل المنحني أن هذا الاحتمال هو أيضاً احتمال وقوع قيم المتغير ع بين العددين – أ ، . أى : ل (٠ ≥ ع > – أ)   (٤)

وذلك لتساوى مساحتي المنطقتين المناظرتين .

في هذا الجدول يعبر الهامش الرأسي (الذى على اليسار أو على اليمين) عن

قيم أ إلى خانة عشرية واحدة ، ويعبر الهامش الأفقي (الذى في أعلى الجدول) عن الخانة العشرية الثانية أى خانة الجزء من مائة . أما الأعداد التي في قلب الجدول فهى احتمالات وقوع المتغير ع في الفترة (٠ ، أ) .

مثال (٤ – ١) :

إذا علم أن توزيع درجات الطلاب في مادة ما هو توزيع معتدل وسطه الحسابي μ = ٦٠ وانحرافه المعيارى σ = ١٦ . فأوجد باستخدام جدول المساحات الاحتمالات الآتية :

أولاً : ل (٨٠ ≥ س > ٦٠)   **ثانياً** : ل (س > ٨٠)

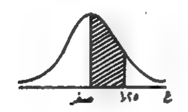
ثالثاً : ل (س ≤ ٨٠)   **رابعاً** : ل (٧٦ ≥ س > ٤٠)

خامساً : ل (٨٥ ≥ س > ٧٥)

الحل :

لإيجاد الاحتمالات المطلوبة من الجدول يجب أن نحول المتغير المعتدل س إلى المتغير المعتدل المعيارى ع بواسطة الصيغة (٢) وهى هنا

$$ع = \frac{س - ٦٠}{١٦}$$



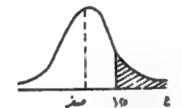
(أولاً) بوضع س = ٦٠ نجد أن $ع = \frac{٦٠ - ٦٠}{١٦}$ = صفر

وبوضع س = ٨٠ نجد أن $ع = \frac{٨٠ - ٦٠}{١٦} = ١,٢٥$

إذن ل (٨٠≥س>٦٠) = ل (١,٢٥ ≥ ع >٠)

= ٠,٣٩٤٤

وذلك من الجدول مباشرة عند العدد ١,٢ الذى في الهامش الرأسي وتحت العدد ٥ الذى في الهامش الأفقي . وهذه النتيجة تعني أن حوالى ٣٩٪ من الطلاب تقع درجاتهم بين ٦٠ ، ٨٠ .



(ثانيا) بوضع س = ٨٠ نجد أن ع = ١,٢٥

∴ ل (س > ٨٠) = ل (ع > ١,٢٥)

= مساحة المنطقة التي على يمين العدد ١,٢٥

= ٠,٥٠٠٠ − ٠,٣٩٤٤

= ٠,١٠٥٦

أى أن حوالى ١١٪ من الطلاب تزيد درجاتهم عن ٨٠ .

(ثالثاً) ل (س ≤ ٨٠) = ل (ع ≤ ١,٢٥)

= مساحة المنطقة التي على يسار العدد ١,٢٥

= ٠,٣٩٤٤ + ٠,٥٠٠٠ = ٠,٨٩٤٤

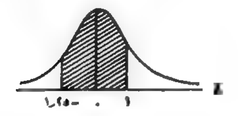


أى أن حوالى ٨٩٪ من الطلاب درجاتهم تساوى أو تقل عن ٨٠ .

يلاحظ أن جواب أى من الاسئلة الثلاثة السابقة يمكن الحصول عليه من جوابى السؤالين الآخرين .

(رابعاً) بوضع س = ٤٠ نجد أن ع = $\frac{٤٠ - ٦٠}{١٦}$ = −١,٢٥

بوضع س = ٧٦ نجد أن ع = $\frac{٧٦ - ٦٠}{١٦}$ = ١



إذن ل (٧٦ ⩾ س > ٤٠) = ل (١ ⩾ ع > − ١,٢٥)

= مساحة المنطقة فوق الفترة (٠ ، ١) + مساحة المنطقة فوق الفترة (− ١,٢٥ ، ٠) = ٠,٣٤١٣ + ٠,٣٩٤٤ = ٠,٧٣٥٧

(خامساً) س = ٧٥ تعطى ع = ٠,٩٣٧٥ = ٠,٩٤ تقريباً

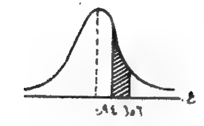
، س = ٨٥ تعطى ع = ١,٥٦٢٥ = ١,٥٦

إذن ل (٨٥ ⩾ س > ٧٥) = ل (١,٥٦ ⩾ ع > ٠,٩٤)

= المساحة فوق الفترة (٠ ، ١,٥٦)

− المساحة فوق الفترة (٠ ، ٠,٩٤)

= ٠,٤٤٠٦ − ٠,٣٢٦٤ = ٠,١١٤٢



مثال (٤ − ٢) : مثال مشهور

للمتغير المعتدل س الذى وسطه الحسابي μ وانحرافه المعيارى σ اثبت أن :

(أولا) احتمال وقوع قيم المتغير بين العددين $\mu \pm \sigma$ هو ٠,٦٨٢٦

(ثانيا) احتمال وقوع قيم المتغير بين العددين $\mu \pm ٢\sigma$ هو ٠,٩٥٤٤

(ثالثا) احتمال وقوع قيم المتغير بين العددين $\mu \pm ٣\sigma$ هو ٠,٩٩٧٤

الحل :

لكى نستخدم الجدول نحول المتغير س إلى المتغير ع بواسطة التعويض :

ع = (س − μ) / σ

أولاً : بوضع : س = $\mu - \sigma$ نجد أن ع = $\frac{(\mu - \sigma) - \mu}{\sigma}$ = −١

وبوضع : س = $\mu + \sigma$ نجد أن ع = $\frac{(\mu + \sigma) - \mu}{\sigma}$ = ١

∴ ل ($\mu + \sigma$ ≥ س > $\mu - \sigma$) = ل (١ ≥ ع > −١)

= ٢ ل (١ ≥ ع > ٠)   من التماثل

= ٢ × ٠,٣٤١٣ = ٠,٦٨٢٦

ثانياً : بالمثل نجد أن :

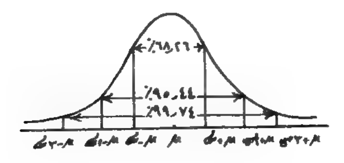
ل ($\mu + ٢\sigma$ ≥ س > $\mu - ٢\sigma$) = ل (٢ ≥ ع > −٢)

= ٢ ل (٢ ≥ ع > ٠) = ٢ × ٠,٤٧٧٢ = ٠,٩٥٤٤

ثالثاً : ل ($\mu + ٣\sigma$ ≥ س > $\mu - ٣\sigma$) = ل (٣ ≥ ع > −٣)

= ٢ ل (٣ ≥ ع > ٠) = ٢ × ٠,٤٩٨٧ = ٠,٩٩٧٤

وتصور هذه النتائج بيانياً كما في الشكل (٤ − ٤) الآتي :



الشكل (٤ − ٤) المساحات أسفل المنحنى المعتدل

مثال (٤ – ٣) : مثال مشهور

بنفس الطريقة نثبت العلاقات الهامة الآتية التي سنحتاج إليها في كثير من التطبيقات الإحصائية ، انظر السؤال (١ – حـ) من تمارين (٤) .

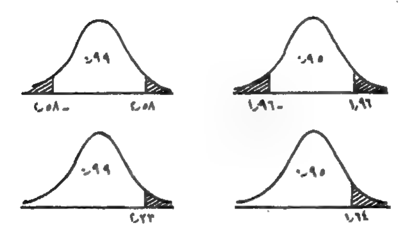
(١) ل (١,٩٦ ⩽ ع < – ١,٩٦) = ٠,٩٥

(٢) ل (٢,٥٨ ⩽ ع < – ٢,٥٨) = ٠,٩٩

(٣) ل (ع > ١,٦٤) = ٠,٠٥

(٤) ل (ع > ٢,٣٣) = ٠,٠١

وتمثل هذه الاحتمالات كما في الشكل (٤ – ٥) الآتي :



الشكل (٤ – ٥) بعض القيم الحرجة للمتغير المعتدل المعيارى

(٤ – ٣) الكشف عن الاعتدالية :

في كثير من الأحيان يبني التحليل الإحصائي لبيانات ناتجة من عينة على أساس افتراض أن المجتمع الذى سحبت منه العينة معتدل ، ولذلك ينبغى أن نتحقق من توفر هذا الافتراض قبل إجراء مثل هذا التحليل .

نفرض أن لدينا توزيعاً تكرارياً ذا فئات لعينة عشوائية وسطها الحسابي $\bar{س}$ وانحرافها المعيارى ع ونريد اختبار ما إذا كانت هذه العينة مأخوذة من مجتمع معتدل .

نتصور مجتمعاً معتدلا له نفس الوسط الحسابي والانحراف المعيارى للتوزيع التكرارى المعلوم أى نأخذ $\mu = \bar{س}$ ، $\sigma =$ ع . إذا كانت العينة مأخوذة من هذا المجتمع فإن احتمال وقوع المتغير المعتدل في فئة مساوية لأى من الفئات التي ينقسم إليها التوزيع التكرارى لا يجب أن يختلف كثيراً عن التكرار النسبي المشاهد في هذه الفئة . وعلى ذلك نقوم بحساب احتمالات وقوع المتغير المعتدل في جميع فئات التوزيع التكرارى مستعينين في ذلك بجدول المساحات . بضرب هذه الاحتمالات (التكرارات النسبية) في حجم العينة نحصل على ما يسمى بالتكرارات النظرية أو المتوقعة المناظرة للتكرارات المشاهدة في العينة . نقارن بين التكرارات المشاهدة والتكرارات المتوقعة لها فإذا كانت قريبة من بعضها بدرجة معقولة أى كانت هناك مطابقة حسنة بين التوزيع التكرارى المشاهد والتوزيع التكرارى النظرى ، جاز لنا أن نعتبر أن المجتمع معتدل ، وإلا فهو ليس كذلك .

إن عملية إيجاد توزيع تكرارى نظرى بالطريقة المذكورة تسمى بعملية **توفيق توزيع معتدل لتوزيع تكرارى معلوم** . وقد سبق أن مرت بنا فكرة التوفيق هذه في حالة كل من توزيع ذى الحدين وتوزيع بواسون . وكما سبق القول ، يعتمد الحكم الموضوعى على حسن المطابقة أو سوئها على أحد الاختبارات الإحصائية مثل اختبار χ^2 الذى سندرسه فيما بعد .

مثال (٤ – ٤) :

وفق التوزيع المعتدل للتوزيع التكرارى الآتي ، وإذا كان هذا التوزيع لعينة عشوائية فاختبر ما إذا كان من الممكن اعتبار أن المجتمع الذى سحبت منه هو مجتمع معتدل .

الفئــــــة	٣١٥-	٣٢٥-	٣٣٥-	٣٤٥-	٣٥٥-	٣٦٥-	٣٧٥-	٣٨٥-	٣٩٥-	٤٠٥-
التكــرار	٦	٦	١١	١٤	١٦	١٥	٨	١٠	٨	٦

الحل :

يتطلب توفيق التوزيع المعتدل أن نوجد الوسط الحسابي والانحراف المعياري للتوزيع التكراري المعطى لاستخدامهما في تقدير الوسط الحسابي والانحراف المعياري للمجتمع . كالمعتاد نجد أن :

الوسط الحسابي $\bar{س} = \frac{1}{ن} \sum ك_ر س_ر = \frac{1}{100}(٦\times٣٢٠ + ٦\times٣٣٠ + \cdots + ٦\times٤١٠) = \frac{٣٦٤٧٠}{١٠٠}$

$= ٣٦٤,٧$

التباين $= ع^2 = \frac{1}{ن-1}\left(\sum ك_ر س_ر^2 - \frac{(\sum ك_ر س_ر)^2}{ن}\right)$

$= \frac{1}{٩٩}\left(٦\times٣٢٠^2 + ٦\times٣٣٠^2 + \cdots + ٦\times٤١٠^2 - \frac{٣٦٤٧٠^2}{١٠٠}\right)$

$= ٧١٢,٨٩$

الانحراف المعياري $= ٢٦,٧$

نريد أن نختبر أن المجتمع الذى سحبت منه العينة هو مجتمع معتدل وسطه الحسابي ٣٦٤,٧ وانحرافه المعياري ٢٦,٧ . نحسب التكرارات المتوقعة كما في الجدول (٤ - ١) الآتي :

الجدول (٤ – ١)

توفيق توزيع معتدل للتوزيع التكراري في المثال (٤ – ٤)

الفئات بقيم س	الفئات بقيم ع	التكرارات النسبية المتوقعة ل ر	التكرارات المتوقعة ق ر = ل ر × ١٠٠	التكرارات المشاهدة ك ر
−∞ − ٣٢٥	−∞ − (−١,٤٩)	٠,٠٦٨١	٦,٨١	٦
٣٢٥ − ٣٣٥	(−١,٤٩) − (−١,١١)	٠,٠٦٥٤	٦,٥٤	٦
٣٣٥ − ٣٤٥	(−١,١١) − (−٠,٧٤)	٠,٠٩٦١	٩,٦١	١١
٣٤٥ − ٣٥٥	(−٠,٧٤) − (−٠,٣٦)	٠,١٢٩٨	١٢,٩٨	١٤
٣٥٥ − ٣٦٥	(−٠,٣٦) − ٠,٠١	٠,١٣٦٦	١٣,٦٦	١٦
٣٦٥ − ٣٧٥	٠,٠١ − ٠,٣٩	٠,١٥٥٧	١٥,٥٧	١٥
٣٧٥ − ٣٨٥	٠,٣٩ − ٠,٧٦	٠,١٢٤٧	١٢,٤٧	٨
٣٨٥ − ٣٩٥	٠,٧٦ − ١,١٣	٠,٠٩٤٤	٩,٤٤	١٠
٣٩٥ − ٤٠٥	١,١٣ − ١,٥١	٠,٠٦٣٧	٦,٣٧	٨
٤٠٥ − ∞	١,٥١ − ∞	٠,٠٦٦٥	٦,٥٥	٦
		١,٠٠٠٠	١٠٠,٠٠	١٠٠

في العمود الأول من هذا الجدول نضع الفئات كما هي معطاة مع تعديل واحد وهو وضع − ∞ بدلا من الحد الأدنى للفئة الأولى و + ∞ بدلا من الحد الأعلى

للفئة الأخيرة ، وذلك لأن التوزيع المعتدل هو توزيع متصل تقع قيمه بين −∞ ، + ∞ وهذا التعديل من شأنه إدخال جميع هذه القيم دون أن يؤثر ذلك على التوزيع التكرارى المعطى .

وفي العمود الثاني نضع حدود الفئات بعد تحويلها إلى قيمها المعيارية بواسطة التعويض $ع = \frac{س - ٣٦٤,٧}{٢٦,٧}$ توطئة لاستخدام جدول المساحات أسفل المنحني المعتدل المعيارى ، فمثلا للفئة الأولى

س = − ∞ تعطى ع = − ∞

، س = ٣٢٥ تعطى $ع = \frac{٣٢٥ - ٣٦٤,٧}{٢٦,٧} = -١,٤٩$

وهكذا بالنسبة لبقية الفئات .

وفي العمود الثالث نضع التكرارات النسبية المتوقعة $ل_ر$ التي تعبر عن احتمالات وقوع قيم المتغير ع في الفئات المناظرة ، وهذه الاحتمالات نوجدها من جدول المساحات . فمثلا للفئتين الأولى والثانية :

ل (−١,٤٩ ≤ ع > −∞) = ٠,٥٠٠٠ − ٠,٤٣١٩ = ٠,٠٦٨١

ل (−١,١١ ≤ ع > −١,٤٩) = ٠,٤٣١٩ − ٠,٣٦٦٥ = ٠,٠٦٥٤

وهكذا بالنسبة لبقية الفئات .

ولما كانت هذه الاحتمالات هى بمثابة التكرارات النسبية في كل فئة فإننا للمقارنة بالتوزيع المعطى نضرب كلا منها في حجم هذا التوزيع وهو هنا ١٠٠ لنحصل على التكرارات المتوقعة $ق_ر$ أى التكرارات التي نتوقعها في حالة كون المجتمع معتدلا . وهذه نضعها في العمود الرابع . أما العمود الخامس فيحمل التكرارات المشاهدة في العينة لتسهيل مقارنتها بالتكرارات المتوقعة .

ومن هذه المقارنة نشعر أن هناك مطابقة حسنة بين التوزيع التكرارى المشاهد والتوزيع التكرارى النظرى إذ لا تختلف التكرارات المشاهدة عن نظائرها المتوقعة في أغلب الفئات إلا قليلا مما يشير إلى أن المجتمع الذى أخذت منه العينة هو على الأرجح مجتمع معتدل ، وإن كان الحكم الموضوعى في ذلك يتطلب استخدام أحد الاختبارات الإحصائية المناسبة . انظر المثال (٦ - ٩) في البند (٦ - ٧ - ٢) .

(٤ - ٤) طريقة بيانية للكشف عن الاعتدالية :

هناك طريقة بيانية تستخدم كاختبار سريع للكشف عن دلالة المنحني التكرارى المشاهد ومدى انحرافه عن الاعتدالية ، ويشترط في هذه الطريقة أن تكون العينة عشوائية وكبيرة الحجم (ن أكبر من ٥٠) . وتبنى فكرة هذه الطريقة على أن التوزيع المعتدل هو توزيع متماثل ذو تفرطح معين وبالتالى فإن أهم ما ينبغى التحقق منه في توزيع تكرارى لعينة هو مدى تماثله ومدى تفرطحه بالنسبة للتوزيع المعتدل .

وكل ما تتطلبه هذه الطريقة هو تكوين توزيع التكرارات المتجمعة المئوية للتوزيع التكرارى المعلوم ثم رسم النقط التي تمثل هذا التوزيع على ورق تقسيم الاحتمالات . فإذا وقعت هذه النقط على وجه التقريب على خط مستقيم يمكن توفيقه بالعين دل ذلك على أن المجتمع هو على الأرجح مجتمع معتدل ، وإلا فهو ليس كذلك . راجع الخاصة (هـ) من البند (٤ - ١) .

ولهذه الطريقة فائدة أخرى ، وهى أنه إذا ظهر لنا أن المجتمع معتدل أو قريب من الاعتدال فإننا نستطيع تقدير الوسط الحسابي والانحراف المعيارى لهذا المجتمع من المستقيم الذى وفقناه كالآتي :

(أ) الوسط الحسابي (الوسيط في الواقع) يقدر بالإحداثي السيني للنقطة التي على الخط المستقيم التي إحداثيها الصادى ٥٠ .

(ب) الانحراف المعيارى يقدر بنصف الفرق بين الإحداثيين السينيين للنقطتين اللتين إحداثياهما الصاديان ١٥,٩ ، ٨٤,١ .

مثال (٤ – ٥) :

استخدم الطريقة البيانية لاختبار اعتدالية المجتمع الذى سحبت منه العينة المذكورة في المثال (٤ – ٤) ، وإذا رأيت أن المجتمع معتدل فأوجد تقديراً لوسطه الحسابي وانحرافه المعيارى .

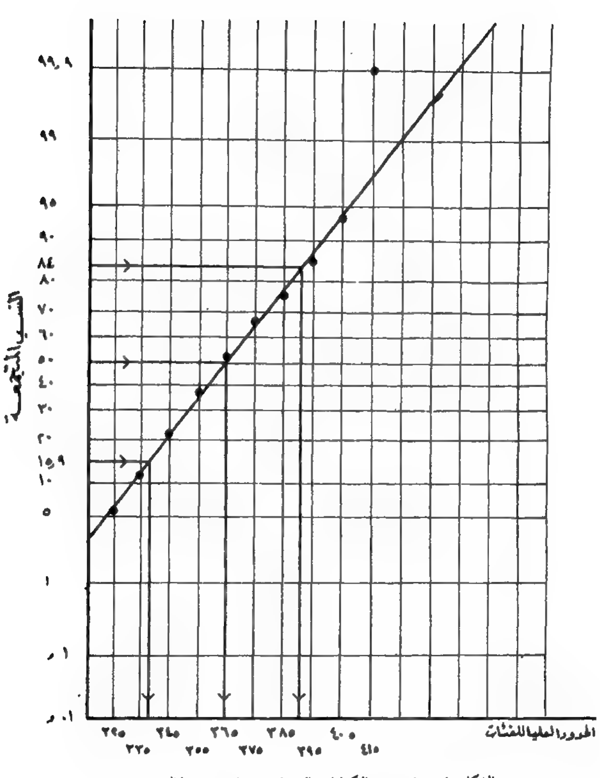
الحل :

نبدأ بإنشاء توزيع التكرارات المتجمعة المئوية كما في الجدول (٤ – ٢) الآتي :

الجدول (٤ – ٢)

الحدود العليا للفئات	التكرار المتجمع	التكرار المتجمع ٪
≥ ٣٢٥	٦	٦
≥ ٣٣٥	١٢	١٢
≥ ٣٤٥	٢٣	٢٣
≥ ٣٥٥	٣٧	٣٧
≥ ٣٦٥	٥٣	٥٣
≥ ٣٧٥	٦٨	٦٨
≥ ٣٨٥	٧٦	٧٦
≥ ٣٩٥	٨٦	٨٦
≥ ٤٠٥	٩٤	٩٤
≥ ٤١٥	١٠٠	١٠٠

نرسم النقط (٣٢٥ ، ٦) ، (١٢،٣٣٥) ، ... ، (٤١٥ ، ١٠٠) على ورق تقسيم الاحتمالات المعتدلة لنحصل على الشكل (٤ – ٦) الآتي :



الشكل (٤ – ٦) توزيع التكرارات النسبية المتجمعة لبيانات المثال (٤ – ٤)

مرسوماً على ورق تقسيم الاحتمالات المعدلة

بالتأمل في هذا الشكل نجد أن النقط تكاد تقع على خط مستقيم مما يشير إلى اعتدالية التوزيع . من الخط المرسوم نقدر الوسط الحسابي والانحراف المعيارى للمجتمع كالآتي :

الوسط الحسابي = ٣٦٥

الانحراف المعيارى = $\frac{١}{٢}$ (٣٩٢ − ٣٣٩) = ٢٦,٥

(٤ – ٥) معالجة عدم اعتدالية التوزيع :

في التحليل الإحصائي للبيانات يتطلب الأمر في بعض الحالات أن يكون المجتمع معتدلا ، فإذا لم يكن المجتمع معتدلا نبحث عن تحويل مناسب يجعله معتدلا أو قريباً من الاعتدال . ومن أكثر التحويلات استخداماً في هذا الصدد التحويل المسمى بالتحويل اللوغاريتمى logarithmic transformation الذى يحول المتغير س الذى لدينا إلى متغير ص حيث ص = لو س . ولا بأس من أخذ اللوغاريتمات العادية أى ذات الأساس ١٠ . وفي كثير من الحالات يفلح هذا التحويل في تعديل التوزيعات التكرارية الملتوية إلى اليمين إلى توزيعات أكثر تماثلا وبالتالى يكون قد عالج إلى حد ما عدم اعتدالية التوزيع . هذا مع ملاحظة أن عدم الاعتدالية لا تترتب عليه نتائج وخيمة إلا إذا كان التوزيع ملتوياً التواء شديداً . على أنه إذا لم يوجد تحويل يصلح لتصحيح الاعتدالية فإننا نلجأ في تحليل البيانات إلى طرق لا تتطلب شرط الاعتدالية وهذه الطرق تسمى بالطرق غير البارامترية non-parametric methods وهى طرق لا تعتمد على فرض توزيعات محددة للمجتمعات أو المتغيرات التي ندرسها . انظر الفصل الرابع عشر .

ملاحظة عن التحويلات :

(١) يستخدم التحويل اللوغاريتمى أيضاً في تحويل نموذج من النوع الضربي مثل $ص = س_{١} س_{٢} س_{٣}$ إلى نموذج من النوع الخطى $صَ = سَ_{١} + سَ_{٢} + سَ_{٣}$

الذى هو أسهل تناولا ، وذلك بوضع ص = لو ص ، س = لو س ... الخ ، وهذا ما نفعله أحياناً في موضوع تحليل الانحدار . كما يستخدم التحويل اللوغاريتمى حين يكون الوسط الحسابي μ للمجتمع مرتبطاً ارتباطاً موجباً بالتباين σ^2 فهو يحول المتغير الذى لدينا إلى متغير آخر يكون فيه هذان الدليلان مستقلين .

(٢) هناك تحويل آخر يسمى بتحويل الجذر التربيعي square-root transformation حيث نضع ص = $\sqrt{س}$. ويستخدم هذا التحويل للبيانات التي تنتج عن العد ويكون توزيعها بواسونيا حيث يكون الدليلان μ ، σ^2 غير مستقلين (إذ نعلم أن $\mu = \sigma^2$) ويفلح هذا التحويل في جعلهما مستقلين . وإذا احتوت البيانات على أصفار يفضل استخدام التحويل ص = $\sqrt{س + \frac{1}{2}}$.

(٣) من التحويلات الشهيرة أيضاً التحويل الزاوى angular transformation حيث نضع ص = $حا^{-1} \sqrt{س}$ ويستخدم حين تكون البيانات مؤلفة من نسب مئوية . وفي توزيع ذى الحدين الذى دليلاه ن ، ح نعلم أن الوسط الحسابي μ = ن ح والتباين σ^2 = ن ح (١ − ح) وبالتالى فإن التباين يكون دالة في الوسط الحسابي . إن التحويل الزاوى يوقف هذه الدالية . إلا أنه حين تكون النسب واقعة بين ٠,٣٠ ، ٠,٧٠ فإن هذا التحويل لا يكون له ضرورة .

(٤) إن التحويلات الثلاثة السابقة تغير العلاقة الدالية بين المتغيرات أى تغير النموذج الرياضى إلى نموذج آخر . غير أن هناك تحويلات لا تفعل هذا وإنما تستهدف تقنين المتغيرات عن طريق :

(أ) استبعاد وحدات القياس وذلك بالتحويل إلى مقياس نسبى لا يعتمد على وحدات القياس ،

(ب) جعل القيم الناتجة عن التحويل فى المجموعات المختلفة تتساوى فى أوساطها الحسابية وفى تبايناتها .

وأشهر هذه التحويلات يأخذ الصورة الآتية المسماه بالصورة المعيارية :

$$ص_ر = \frac{س_ر - \bar{س}}{ع}$$

حيث $\bar{س}$ متوسط قيم $س$ ، $ع$ انحرافها المعيارى .

ونظرا لأن هذا المقياس نسبى فإن القيم الصادية الناتجة تكون خالية من أى وحدة قياس ، كما أن هذا التحويل إذا أجرى على قيم المتغير فى أى مجموعة فإنه يحول هذه القيم إلى قيم متوسطها يساوى صفرا وتباينها يساوى الواحد الصحيح .

تمارين (٤)

(١) للتوزيع المعتدل المعيارى وباستخدام جدول المساحات

(أ) أوجد كلا من الاحتمالات الآتية :

ل (١,٥ ⩽ ع < ٠) ، ل (ع ⩽ ١,٦٥) ، ل (ع < − ٠,٧٤)

، ل (١,٢٥ ⩽ ع < − ٢,٠٥) .

(ب) أوجد قيمتي أ ، ب بحيث ل (ع < ب) = ٠,٩٠ ،

ل (ع < أ) = ٠,٤٨

(جـ) أثبت أن ل (١,٩٦ ⩽ ع < − ١,٩٦) = ٠,٩٥ ، ل (ع ⩽ ١,٦٤)

= ٠,٠٥

، ل (٢,٥٨ ⩽ ع < − ٢,٥٨) = ٠,٩٩ ، ل (ع ⩽ ٢,٣٣) = ٠,٠١

(٢) للتوزيع المعتدل الذى وسطه الحسابي ٥٠ وانحرافه المعيارى ٥ أوجد كلا من

ل (٦٠,٥ ⩽ س < ٣٠,٥) ، ل (٥٥ ⩽ س < ٤٥) .

(٣) في مجتمع معين المعروف أن نسب الذكاء تتوزع توزيعاً معتدلا وسطه الحسابي ١٠٤,٦ وانحرافه المعيارى ١٥,٣ .

(أ) أوجد نسبة الأفراد الذين تقع نسب ذكائهم بين ٩٠ ، ١٢٠

(ب) أوجد نسبة الأفراد الذين تزيد نسب ذكائهم عن ١١٠

(٤) التوزيع الآتي هو التوزيع التكرارى لعينة عشوائية مأخوذة من مجتمع ما .

٢٧,٥−	٣٢,٥ −	٣٧,٥−	٤٢,٥−	٤٧,٥−	٥٢,٥−	٥٧,٥−	٦٢,٥−
٥	٦	١٣	٨	١٦	١٤	٧	١

(أ) أثبت أن الوسط الحسابي يساوى ٤٧,٠٧١ وأن الانحراف المعيارى يساوى ٨,٩٤٨ .

(ب) وفق توزيعاً معتدلا لهذا التوزيع واذكر رأيك فيما إذا كان بالامكان اعتبار أن المجتمع معتدل .

(جـ) استخدم الطريقة البيانية لاختبار اعتدالية المجتمع .

(٤ – ٦) تقريب توزيع ذى الحدين بتوزيع معتدل :

فى البند (٣ – ٤ – ١) رأينا أنه إذا كان س متغيرا عشوائيا ذا توزيع ذى حدين دليلاه ن ، ح يجوز تقريب هذا التوزيع بتوزيع بواسون متوسطه يساوى متوسط توزيع ذى الحدين بشرط أن يكون حجم العينة ن كبيرا وأن يكون الاحتمال ح صغيرا حيث يكون التوزيع ملتويا إلى اليمين . يجوز تحت شروط أخرى تقريب توزيع ذى الحدين : حد (ن ، ح) بتوزيع معتدل متوسطه μ = ن ح أى يساوى متوسط توزيع ذى الحدين ، وتباينه σ^2 = ن ح ك أى يساوى تباين ذى الحدين بشرط أن تكون ن كبيرة وأن تكون ح قريبة من العدد $\frac{1}{2}$ حيث يكون التوزيع متماثلا بالتقريب . تحت هذين الشرطين يقترب توزيع الإحصاءة

$$ع = \frac{س - ن\,ح}{\sqrt{ن\,ح\,ك}} \qquad \text{حيث } ك = ١ - ح$$

من التوزيع المعتدل المعيارى : مع (١٦٠) كلما زاد العدد ن .

ومن الناحية العملية وجد أن هذا التقريب يكون جيداً أى يمكن التجاوز عن الخطأ الناشيء عنه إذا توفر أحد الشرطين الآتيين :

(أ) إذا كانت كل من ن ح و ن ك أكبر من خمسة

أو (ب) إذا كانت ن ⩾ ١٠ ومعامل الالتواء أصغر من ٠,٢ .

ونظرا لأن توزيع ذى الحدين هو توزيع لمتغير وثاب بينما التوزيع المعتدل هو توزيع لمتغير متصل فإننا لعلاج ذلك فى عملية التقريب يجب أن تعتبر أن كل قيمة س من قيم المتغير ذى الحدين ممتدة نصف وحدة من اليسار ونصف وحدة من اليمين فمثلا العدد س = ٤ نعتبر أنه الفترة (٣,٥ ، ٤,٥) ، والعدد س= ٤,٧ نعتبر أنه الفترة (٤,٦٥ ، ٤,٧٥) ، وإذا أردنا مثلا ايجاد الإحتمال ل (١٠ ⩾ س ⩾ ٣) لتوزيع ذى حدين فإننا نحسب الاحتمال التقريبي ل(١٠,٥ ⩾س⩾٢,٥) باستخدام جداول التوزيع المعتدل .

مثال (٤ – ٦)

ألقيت حجرة نرد منتظمة عشوائيا ١٠ مرات . أوجد احتمال الحصول على الصورة في ٣ أو ٤ أو ٥ أو ٦ مرات .

الحـــل :

نظرا لأن حجرة النرد منتظمة والرمية عشوائية فإن توزيع عدد مرات ظهور الصورة هو توزيع ذو حدين : حد (١٠ ، ٠,٥) ودالة كتلة الاحتمال تكون كالآتى :

$$\text{د (س)} = \text{ق}_{\text{س}}^{١٠} (٠,٥)^{\text{س}} (٠,٥)^{١٠-\text{س}} = (٠,٥)^{١٠} \text{ق}_{\text{س}}^{١٠}$$

حيث س = ٠ ، ١ ، ٢ ، ... ، ١٠

∴ ل (س = ٣ أو ٤ أو ٥ أو ٦) = د (٣) + د(٤) + د(٥) + د(٦)

$$= (٠,٥)^{١٠} \left(ق_{٣}^{١٠} + ق_{٤}^{١٠} + ق_{٥}^{١٠} + ق_{٦}^{١٠} \right)$$

$$= ٠,٧٧٣٤$$

الحل التقريبى :

بما أن الشرط (ب) متوفر إذ أن ن= ١٠≤١٠ ومعامل الالتواء= صفر<٠,٢ (مع ملاحظة أن التوزيع متماثل تماما لأن ح = $\frac{١}{٢}$) يمكن التقريب بتوزيع معتدل متوسطه μ= ن ح = ١٠×٠,٥= ٥ وتباينه $\sigma^{٢}$ = ن ح ك = ١٠×٠,٥×٠,٥= ٢,٥ واذن انحرافه المعيارى يساوى $\sqrt{٢,٥}$ = ١,٥٨. ويكون للمتغير

$$ع = \frac{س - ٥}{١,٥٨}$$

توزيع معتدل معيارى على وجه التقريب . وهنا نعتبر أن العدد ٣ هو الفترة (٢,٥ ، ٣,٥) ، وأن العدد ٤ هو الفترة (٣,٥ ، ٤,٥) وهكذا ... ويكون المطلوب إيجاد احتمال ل(٦,٥ >س>٢,٥) فى التوزيع المعتدل مع (٥ ، ١,٥٨)

بوضع س= ٢,٥ نجد أن ع = $\frac{٢,٥ - ٥}{١,٥٨}$ = ــ ١,٥٨٢

بوضع س= ٦,٥ نجد أن ع = $\frac{٦,٥ - ٥}{١,٥٨}$ = ٠,٩٥

ل(٦,٥>س>٢,٥) = ل (٠,٩٥>ع> − ١,٥٨٢) = ٠,٧٧١٨

ومن الواضح أن هذه القيمة قريبة جدا من القيمة المضبوطة ٠,٧٧٣٤

(ثانيا) التوزيع المعتدل اللوغاريتمى

إن التوزيع المعتدل اللوغاريتمى هو مثال آخر لنماذج الاحتمال المتصلة ، وقد سمى كذلك لأن التحويل ص = لو س يحول المتغير الذى يصفه هذا التوزيع إلى متغير ذى توزيع معتدل . ومن الظواهر التى يصلح لها هذا النموذج بعض الظواهر الجيولوجية كتلك المتعلقة بأوزان وأعداد بعض أنواع الصخور الرسوبية ، وبعض الظواهر الاقتصادية كدخول الأفراد وخاصة الدخول ذوات القيم الصغيرة .

ويعرّف هذا التوزيع بواسطة دالة كثافة الاحتمال الآتية :

$$\text{د}(\text{س}) = \frac{1}{\text{س}\sqrt{2\,\text{ط}}\;\theta^{2}}\;\text{هـ}^{-\frac{1}{2\theta^{2}}(\text{لو س}-\gamma)^{2}} \qquad \text{س} > 0 \qquad (4)$$

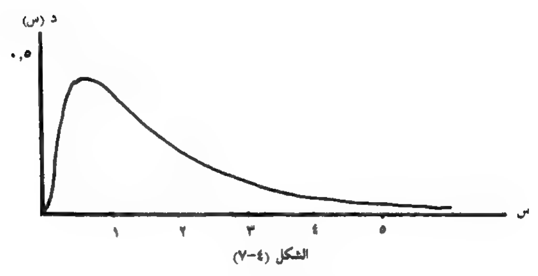
حيث اللوغاريتم للأساس هـ وحيث γ ، θ بارامتران إذا علمت قيمتاهما تحدد التوزيع تحديدا تاما .

(٤ - ٦) بعض خصائص التوزيع

(أ) المنحنى الممثل للدالة (٤) هو منحنى ذو قمة واحدة وملتوى إلى اليمين ، ويختلف شكله باختلاف قيمتى البارامترين γ ، θ ، فمثلا يأخذ الشكل المبين بالشكل (٤-٧) حين تأخذ γ القيمة صفر وتأخذ θ القيمة ١ .

(ب) التحويل ص = لو س حيث اللوغاريتم للأساس هـ يحول المتغير س إلى متغير ص له توزيع معتدل وسطه الحسابى γ وانحرافه المعيارى θ ، وبالتالى يكون المتغير

$$\text{ع} = \frac{\text{لو س} - \gamma}{\theta}$$



الشكل (٤-٧)

المنحنى المعتدل اللوغاريتمي: μ=٠ وθ = ١

هو متغير معتدل معيارى (وسطه الحسابى صفر وتباينه ١) . وبناء على ذلك يمكن استخدام جدول المساحات أسفل المنحنى المعتدل المعيارى ، وهو الجدول (٦) بملحق هذا الكتاب ، فى إيجاد الاحتمالات والقيم الحرجة للمتغير س كما فى المثالين الآتيين :

مثال (٤-٦)

إذا كان س متغيرا معتدلا لوغاريتميا دليلاه μ = ١ ، θ = ٢ فأوجد الاحتمال ل(س > ٣٥,٥) .

الحل :

بوضع ص = لو(س) يكون للمتغير ع = $\frac{\text{لو س} - ١}{٢}$ توزيع معتدل معيارى.

ل (س > ٣٥,٥) = ل (لو س > لو ٣٥,٥) = ل $\left(\frac{\text{لو س} - ١}{٢} > \frac{\text{لو ٣٥,٥} - ١}{٢}\right)$

= ل $\left(ع > \frac{٣,٥٦٩٥ - ١}{٢}\right)$ = ل (ع>١,٢٨٤٨)

= ٠,٨٩٩٧ . (من جدول المساحات أسفل المنحنى المعتدل)

مثال (٤-٧)

إذا كان س متغيرا معتدلا لوغاريتميا دليلاه γ = ٢ ، θ= ٠,٥ . فأوجد قيمة ا بحيث ل (س < ا) = ٠,٩٠

الحـل :

ل (س < ا) =ل (لو س < لو ا) = ل $\left(\frac{\text{لو س} - ٢}{٠,٥} < \frac{\text{لو ا} - ٢}{٠,٥}\right)$

= ل (ع < ع) = ٠,٩٠

من جدول المساحات نجد أن ع = ١,٢٨ ∴ $\frac{\text{لوا}-٢}{٠,٥}$ = ١,٢٨

∴ لوا = ٢,٦٤ ∴ ا = ١٤,٠١٣

الفصل الخامس

توزيعات خاصة

SPECIAL DISTRIBUTIONS

كل من التوزيعات الثلاثة الآتية هو توزيع احتمال لمتغيرعشوائى متصل مركب بطريقة معينة من عدد من المتغيرات العشوائية المعتدلة . وهذه التوزيعات لا تستخدم كنماذج احتمال للمجتمعات كما هو الحال في التوزيعات التي عرضت في الفصلين السابقين ، وإنما تبرز من خلال التحليل الإحصائي للعينات وتبني عليها اختبارات إحصائية ذات أهمية قصوى في عملية الاستدلال الإحصائي كما سنرى في الفصل التالى . ومن الناحية التطبيقية يهمنا بصفة خاصة في دراسة هذه التوزيعات أمرين هما :

(١) الشكل الهندسي العام لكل توزيع .

(٢) كيفية استخدام الجداول لإيجاد الاحتمالات والقيم الحرجة .

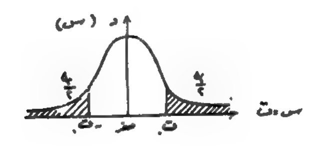
(٥-١) توزيع ت : STUDENT t- DISTRIBUTION

يقال لمتغير عشوائي س إن له توزيع ت إذا وفقط إذا كانت دالة كثافة احتماله معرفة بالقاعدة :

$$د(س) = \frac{[\frac{1}{2}(ن+1)]!}{\sqrt{ن ط}\;[\frac{1}{2}(ن-2)]!}\left(1+\frac{س^2}{ن}\right)^{-\frac{1}{2}(ن+1)} \qquad (1)$$

حيث $\infty > س > -\infty$ ، ن = ١ ، ٢ ، ٣ ...

وهذا التوزيع له دليل واحد هو ن يسمى بعدد درجات الحرية للتوزيع ، ولهذا فالتوزيع يتحدد تماماً إذا علمت قيمة ن .



الشكل (٥-١) منحني توزيع المتغير ت

والمنحني الممثل للدالة ص = د (س) المعرفة في (١) هو منحني ذو قمة واحدة ومتماثل حول المستقيم س = صفر .

الوسط الحسابي للتوزيع : μ = صفر (٢)

، تباين التوزيع : $\sigma^2 = \frac{ن}{ن-2}$ حيث ن > ٢ (٣)

وجدير بالذكر أن توزيع ت يقترب من التوزيع المعتدل المعيارى كلما اقتربت درجات الحرية ن من اللانهاية .

جدول القيم الحرجة :

الجدول (٧) بملحق هذا الكتاب يعطى القيمة الحرجة $ت_\alpha$ critical value للمتغير ت وهى تلك القيمة الموجبة التي تجعل مساحة المنطقة أسفل المنحني وخارج الفترة

(−ت. ، ت.) مساوية لقيمة معينة α عند درجة الحرية ن . أي أن العدد α يعبر عن المساحة عند ذيلي المنحنى ($\frac{\alpha}{2}$ عند كل ذيل) وبالتالى فهو يعبر عن احتمال وقوع قيم المتغير ت خارج الفترة المذكورة . (انظر الشكل ٥−١) . ونكتب هذا رمزياً كالآتي :

$$ل \left(\left| ت \right| > ت \right) = \alpha \qquad (٤)$$

فى هذا الجدول كتبت درجات الحرية في كل من العمودين الهامشيين الواقعين في يمين ويسار الجدول ، وكتبت الاحتمالات α = ٠,٠٠١ ، ٠,٠١ ، ٠,٠٢ ، ٠,٠٥ ، ٠,١ ، ٠,٢ ، ٠,٤ ، ٠,٥ ، ٠,٩ في الهامش الأفقى الذى بأعلى الجدول . أما القيم الحرجة فمدونة في الخلايا التي بقلب الجدول . وحين تكون درجة الحرية ن معلومة لنا ، نستطيع استخراج مايلى :

(١) القيمة الحرجة ت إذا أعطيت قيمة الاحتمال α . ونجد تلك القيمة عند نقطة التقاء الصف الذى به درجة الحرية ن والعمود الذى به قيمة α .

(٢) قيمة الاحتمال α إذا أعطيت القيمة الحرجة . ولإيجاد تلك القيمة نبحث في الصف الذى به درجة الحرية عن القيمة الحرجة المعطاة فتكون α هى العدد الذى يعلو العمود الذى به هذه القيمة .

مثال (٥−١) :

ليكن س متغيراً له توزيع ت بدرجات حرية عددها ٨ . من الجدول نجد مايلى :

(أ) إذا كانت α = ل (|ت| > ت.)=٠,٠٥ فإن القيمة الحرجة ت =٢,٣٠٦

(ب) الاحتمال ل (|ت| >٢,٨٩٦) = ٠,٠٢

(حـ) الاحتمال ل (ت > ٢,٨٩٦) = ٠,٠١ مساحة الذيل الأيمن فقط .

تقليـــد :

نكتب $ت_{\alpha[ن]}$ (٥)

للتعبير عن القيمة الحرجة تبِ في توزيع ت عند درجة الحرية ن بحيث يكون مجموع مساحتي المنطقتين عند الذيلين يساوى α ، فمثلا :

$ت_{٠,٠٥[٧]}$ = ٢,٣٦٥ ، $ت_{٠,٠١[٣]}$ = ٥,٨٤١ ، $ت_{٠,٠٥[\infty]}$ = ١,٩٦

(٥-٢) توزيع χ^2 : THE CHI-SQUARE DISTRIBUTION

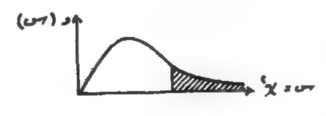
يقال لمتغير عشوائى س إن له توزيع χ^2 (تنطق كاى تربيع) إذا وفقط إذا كانت دالة كثافة احتماله معرفة بالقاعدة :

$$د(س) = \frac{١}{٢^{\frac{ن}{٢}} \Gamma(\frac{ن}{٢})} س^{\frac{ن}{٢}-١} هـ^{-\frac{س}{٢}} \qquad (٦)$$

حيث ∞ > س >٠ ، ن = ١ ، ٢ ، ٣ ، ....

وهذا التوزيع له دليل واحد هو ن يسمى بعدد درجات الحرية للتوزيع ، ولهذا فالتوزيع يتحدد تماماً إذا علمت قيمة ن .

والمنحني الممثل للدالة ص = د (س) المعرفة في (٦) وعندما ن >٢ يكون منحنى ذو قمة واحدة وملتوى إلى اليمين .



الشكل (٥-٢) منحنى توزيع المتغير χ^2 ، (ن > ٢)

الوسط الحسابي للتوزيع : $\mu = ن$ (٧)

، تباين التوزيع : $\sigma^2 = ٢ ن$ (٨)

جدول القيم الحرجة :

الجدول (٨) بملحق هذا الكتاب يعطى القيمة الحرجة χ^2_α للمتغير χ^2 وهي القيمة التي تجعل مساحة المنطقة أسفل المنحنى وفوق الفترة $(\chi^2_\alpha ، \infty)$ مساوية لقيمة معينة α عند درجة الحرية ن . أى أن العدد α يعبر عن المساحة عند الذيل الملتوى،وبالتالى فهو يعبر عن احتمال وقوع قيم المتغير χ^2 على يمين العدد χ^2_α (انظر الشكل ٥-٢) . ونكتب هذا رمزياً كالآتى :

$$ل (\chi^2 < \chi^2_\alpha) = \alpha \quad (٩)$$

فى هذا الجدول كتبت درجات الحرية والاحتمالات α والقيم الحرجة بنفس الطريقة التى كتبت بها في جدول ت ، غير أن α تأخذ القيم ٠,٠٠٥ ، ٠,٠١، ٠,٠٢٥ ، ٠,٠٥ ، ٠,١ ، ٠,٥ ، ٠,٩ ، ٠,٩٧٥ ، ٠,٩٩٥ .

وحين تكون درجة الحرية ن معلومة لنا،نستطيع من الجدول استخراج القيمة الحرجة χ^2_α إذا أعطيت قيمة الاحتمال αأو استخراج الاحتمالα بمعلومة القيمة الحرجة χ^2_α .

مثال (٥-٢) :

ليكن س متغيراً له توزيع χ^2 بدرجات حرية عددها ١٠ . من الجدول نجد مايلى :

(أ) إذا كانت ل $(\chi^2 < \chi^2_\alpha)$=٠,١ فإن القيمة الحرجة χ^2_α = ١٥,٩٨٧

(ب) الاحتمال ل ($\chi^2 <$ ١٨,٣١) = ٠,٠٥ .

(حـ) الاحتمال ل (١٨,٣١ > $\chi^2 >$ ٤,٨٧)=٠,٩٠ - ٠,٠٥=٠,٨٥

تقليـد :

نكتب $\chi^{٢}_{\alpha[ن]}$ (١٠)

للتعبير عن القيمة الحرجة $\chi^{٢}$ في توزيع $\chi^{٢}$ عند درجة الحرية ن بحيث تكون مساحة المنطقة التي إلى يمينها مساوية للعدد α ، :

$$\chi^{٢}_{٠,٠٥[٥]} = ٤,٣٥١ \text{ ، } \chi^{٢}_{٠,٠٥[٥]} = ١١,٠٧٠ \text{ ، } \chi^{٢}_{٠,٠١[٦٠]} = ٨٨,٣٧٩$$

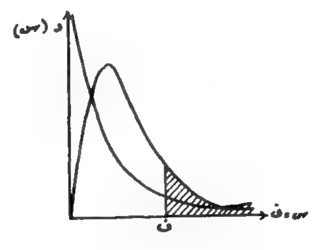
(٥-٣) توزيع ف : THE F-DISTRIBUTION

يقال لمتغير عشوائي س إن له توزيع ف إذا وفقط إذا كانت دالة كثافة احتماله معرفة بالقاعدة :

$$\text{د}(\text{س}) = \frac{\left(\frac{١}{٢}(\text{م}+\text{ن}-٢)\right)! \left(\frac{\text{م}}{\text{ن}}\right)^{\frac{\text{ن}}{٢}}}{\left(\frac{١}{٢}(\text{م}-٢)\right)! \left(\frac{١}{٢}(\text{ن}-٢)\right)!} \times \frac{\text{س}^{\frac{\text{ن}}{٢}-١}}{\left(١+\frac{\text{م}}{\text{ن}}\right)^{\frac{١}{٢}(\text{م}+\text{ن})}} \quad (١١)$$

حيث ∞ > س >٠ ، م ، ن عددان صحيحان موجبان .

وهذا التوزيع له دليلان هما م ، ن يسميان بعددى درجات الحرية ، ويتحدد التوزيع تماماً إذا علمت قيمتي هذين الدليلين .



الشكل (٥-٣) منحنى توزيع المتغير ف

والمنحني الممثل للدالة ص = د (س) المعرفة في (١١) يختلف شكله بحسب قيمتي م ، ن فهو يأخذ الشكل ⌣ إذا كانت م ، ن صغيرتين جداً إلا أنه يصبح محدباً وملتوياً إلتواء شديداً إلى اليمين كلما زادت قيمتا م ، ن .

$$\text{الوسط الحسابي للتوزيع} = \mu = \frac{ن}{ن-2} \qquad \text{حيث } ن > 2 \qquad (12)$$

جدول القيم الحرجة :

الجدول (٩) بملحق هذا الكتاب يعطي القيمة الحرجة ف. للمتغير ف وهي تلك القيمة التي تجعل مساحة المنطقة أسفل المنحني وفوق الفترة (ف. ، ∞) مساوية لقيمة معينة α عند درجتي الحرية م ، ن . أى أن العدد α يعبر عن المساحة عند الذيل الملتوى وبالتالى فهو يعبر عن احتمال وقوع قيم ف على يمين العدد ف . (انظر الشكل ٥-٣) ونكتب هذا رمزياً كالآتي :

$$ل(ف > ف.) = \alpha \qquad (13)$$

ويختلف تركيب هذا الجدول عن جدولى ت وχ^2 إذ يحمل الهامش الأفقى العلوى درجات الحرية م ويحمل كل من العمودين الهامشيين على جانبى الجدول درجات الحرية ن وعند كل من هذه الدرجات وضعت ثلاث قيم للعدد α هى ٠,٠٥ ، ٠,٠٢٥ ، ٠,٠١ وعند تقاطع كل عمود م مع كل صف ن توجد ٣ قيم حرجة تناظر تلك القيم الثلاث .

وحين تكون درجتا الحرية م ، ن معلومتين نستطيع من الجدول استخراج القيمة الحرجة ف. بمعلومية الاحتمال α أو استخراج قيمة الاحتمال α بمعلومية القيمة الحرجة .

مثال (٥–٣) :

ليكن س متغيراً له توزيع ف بدرجتي حرية ٧ ، ٩ .

(ا) إذا كانت ل (ف >ف.) = ٠,٠٥ فإن ف. = ٣,٢٩

(ب) إذا كانت ل (ف >ف.) = ٠,٠١ فإن ف. = ٥,٦١

(حـ) الاحتمال ل (ف >٤,٢) = ٠,٠٢٥

ملاحظة :

لا يعطى الجدول القيم الحرجة إلا عند ثلاث قيم للاحتمال α هي ٠,٠٥ ، ٠,٠٢٥ ، ٠,٠١ ولكن يمكننا أيضاً إيجاد تلك القيم عند الاحتمالات المكملة ٠,٩٥ ، ٠,٩٧٥ ، ٠,٩٩ باستخدام النظرية الآتية :

« إذا كان س متغيراً عشوائياً له توزيع ف بدرجتي حرية م ، ن فإن المتغير $\frac{1}{س}$ يكون له توزيع ف بدرجتي حرية ن ، م » . ويمكن أن نكتب هذه النظرية كالآتى :

$$ف_{\alpha[\text{م، ن}]} = \frac{1}{ف_{(1-\alpha)[\text{ن، م}]}}$$

مثال (٥–٤) :

ليكن س متغيراً له توزيع ف بدرجتي حرية ٥ ، ٩ أوجد قيمة ف. بحيث ل(ف > ف.) = ٠,٩٥

الحـــل :

نلاحظ أن الاحتمال ٠,٩٥ ليس له وجود بالجدول أما الاحتمال المكمل ٠,٠٥ فموجود به . المتباينة ف > ف. تكافئ المتباينة $\frac{1}{ف} < \frac{1}{ف.}$ مع ملاحظة أن ف ، ف. موجبان .

إذن ل(ف > ف.) = ل ($\frac{1}{ف}$ < $\frac{1}{ف.}$) = ٠,٩٥ فرضاً .

إذن ل ($\frac{1}{ف}$ ⩽ $\frac{1}{ف.}$) = ٠,٠٥

من النظرية يكون لدينا متغير له توزيع ف بدرجتي حرية ٩ ، ٥

من الجدول نجد أن $\frac{1}{ف.}$ = ٤,٧٧

إذن ف. = ١ ÷ ٤,٧٧ = ٠,٢٠٩٦

الشكل (٥-٤)

تقليد :

نكتب (١٤)

$$ف_{\alpha[م ، ن]}$$

للتعبير عن القيمة الحرجة في توزيع ف عند درجتي الحرية م ، ن بحيث تكون مساحة المنطقة على يمينها مساوية للعدد α ، فمثلا :

$ف_{٠,٠٥[٥ ، ٨]}$ = ٣,٦٩ ، $ف_{٠,٠٥[٨ ، ٥]}$ = ٤,٨٢ ، $ف_{٠,٠١[١٥ ، ٤]}$ = ١٤,٢

لاحظ أن العدد الأول م يحدد العمود والعدد الثاني ن يحدد الصف .

تمارين (٥)

(١) أوجد كلا من القيم الحرجة الآتية :

(أ) $ت_{٠,٠٥[١٢]}$ ، $ت_{٠,٠٢[٤٠]}$ ، $ت_{٠,٠١[١٦]}$ ، $ت_{٠,٠٥[\infty]}$.

(ب) $\chi^2_{٠,٠٥[٢٠]}$ ، $\chi^2_{٠,٠١[٣٥]}$ ، $\chi^2_{٠,١[٣٦]}$ ، $\chi^2_{٠,٠٥[١]}$.

(جـ) $ف_{٠,٠٥[٨،٦]}$ ، $ف_{٠,٠٥[٦،٨]}$ ، $ف_{٠,٠١[١٠،٤٠]}$ ، $ف_{٠,٠٥[١٤،٣٠]}$.

(٢) أوجد كلا من الاحتمالات الآتية :

(ا) ل (|ت| > ٢,٢٠١) حيث ن = ١١

، ل (ت > ٢,٢٠١) حيث ن = ١١

(ب) ل (χ^2 > ٥٢,١٩) حيث ن = ٣١

، ل (χ^2 > ١٩,٣٤) حيث ن = ٢٠

(جـ) ل (ف > ١٠,٧) حيث م = ٦ ، ن = ٥

، ل (ف > ٨,٥) حيث م = ٣٠ ، ن = ٤

الفصل السادس

نظرية العينات

THEORY OF SAMPLING

(٦-١) نظرية العينات :

تبحث نظرية العينات في العلاقات بين المجتمعات والعينات المأخوذة من هذه المجتمعات ، وقد حوت من النظريات والتوزيعات والأساليب ما يمكننا من تقدير أدلة المجتمعات أو مقارنتها أو إصدار قرارات أو تنبؤات بشأنها عن طريق دراسة وتحليل عينات مأخوذة منها مما يدخل تحت موضوع **الاستدلال الإحصائي** .
STATISTICAL INFERENCE .

ويقصد بالاستدلال الإحصائي أى إجراء يستخدم نظرية الاحتمال في إصدار قرارات عن مجتمع أو عدة مجتمعات عن طريق عينات مأخوذة منها مع تحديد درجة الثقة في هذه القرارات . ولعملية الاستدلال الإحصائي مجالان رئيسيان هما مجال تقدير بارامترات المجتمعات ومجال اختبار الفروض الإحصائية ، على أنه من الشروط الأساسية في هذه العملية أن تكون العينات عشوائية لأن جميع نظريات الاحتمال التي تعتمد عليها مؤسسة على فرض العشوائية . كما أنه كلما كبر حجم العينة كلما كان الاستدلال أكثر دقة .

ومن بين المسائل التي تبرز في الأبحاث التطبيقية وتحتاج إلى عملية الاستدلال الإحصائي ما يلي :

(أ) اختبار صواب أو خطأ فروض مطروحة عن أدلة المجتمعات .

(ب) تقدير متوسطات وتباينات المجتمعات وغيرها من الأدلة .

(جـ) الكشف عما إذا كانت الفروق المشاهدة في العينات هي فروق راجعة إلى الصدفة أو تقلبات العينات أو هي فروق جوهرية تدل على وجود اختلاف حقيقي بين المجتمعات التي أخذت منها هذه العينات .

(د) الكشف عن تأثير واحد أو أكثر من العوامل أو المعالجات على متغير ما أو ظاهرة معينة .

(هـ) تقدير درجة ونوع العلاقة أو الارتباط بين المتغيرات وإصدار تنبؤات عنها .

وسوف نتدارس هذه المسائل وغيرها في هذا الفصل وما يليه من فصول ، بعد تقديم بضعة مفاهيم ونظريات تتطلبها الدراسة الواعية لتلك المسائل .

(٦-٢) توزيعات المعاينة :

فى نظرية العينات نميز بين ثلاثة أنواع من التوزيعات .

(١) توزيع المجتمع Population Distribution

(٢) توزيع العينة Sample Distribution

(٣) توزيع المعاينة Sampling Distribution

ولبيان الفرق بين هذه التوزيعات نفرض أننا نرغب في إيجاد الوسط الحسابي لدخل العائلة في مجتمع مدينة مؤلفة من ٢٠٠٠ عائلة . إذا أمكننا معرفة دخول جميع عائلات المدينة ووضعنا هذه الدخول في توزيع تكرارى فإن هذا التوزيع يسمى بتوزيع المجتمع للدخول . ونستطيع بالطبع أن نحصل مباشرة على **الوسط الحسابى الحقيقي** μ لدخل العائلة في مجتمع المدينة ، أو على أى دليل آخر يخص هذا المجتمع .

أما إذا اخترنا عينة من مجتمع هذه المدينة فإن التوزيع التكرارى لدخول العائلات في هذه العينة يسمى **بتوزيع العينة** . والوسط الحسابي $\bar{س}$ لهذا التوزيع لا يكون عادة مساوياً للوسط الحسابي الحقيقي μ للمجتمع ، إلا أننا قد نأخذ هذا الوسط الحسابي تحت شروط معينة ، كتقدير للوسط الحسابي الحقيقي للمجتمع . وبطبيعة الحال تختلف الأوساط الحسابية للعينات المأخوذة من نفس المجتمع حتي ولو كانت من نفس الحجم .

أما إذا فرضنا أننا حصلنا من المجتمع على جميع العينات التي من **نفس الحجم** ن وأوجدنا الوسط الحسابي لكل من هذه العينات ثم وضعنا هذه الأوساط في توزيع تكرارى فإن هذا التوزيع يسمى حينئذ **بتوزيع المعاينة** للوسط الحسابي (أو للأوساط الحسابية) للعينات ذوات الحجم ن . ويمكننا أن نحسب الوسط الحسابي والانحراف المعيارى ومعامل الالتواء ... لهذا التوزيع . وبالمثل يمكن أن نتصور توزيع المعاينة للتباين للعينات ذوات الحجم ن أو توزيع المعاينة لأى مقياس آخر .

في هذا المثال يستحيل عملياً إيجاد توزيع المعاينة للوسط الحسابي بالطريقة المذكورة لأن عدد العينات الممكنة هو عدد فلكى نعجز عن الحصول عليه . وفي المعتاد نحصل على توزيعات المعاينة بطرق رياضية ، إلا أنه لتوضيح مفهوم توزيع المعاينة نضرب المثال البسيط الآتي .

مثال (٦ — ١) :

يتألف مجتمع من ٦ أرانب أوزانها بالأوقيات ١١ ، ١٦ ، ١٢ ، ١٥ ، ١٦ ، ١٤ .

(أولا) أوجد توزيع المعاينة للوسط الحسابي للعينات ذوات الحجم ٢ (اعتبر أن المعاينة مع الإرجاع) .

(ثانيا) أحسب الوسط الحسابى والتباين للتوزيع الناتج وقارنهما بالوسط الحسابي والتباين للمجتمع .

الحل :

(اولا) بما أن المعاينة مع الإرجاع أى مع رد كل عدد يؤخذ إلى المجتمع قبل أخذ عدد آخر فإن عدد العينات ذوات الحجم ٢ هو ٦ × ٦ = ٣٦ لأن أى عدد من الأعداد الستة يمكن أن يقترن (في عينة حجمها ٢) بأى عدد من الأعداد الستة (بما في ذلك نفسه) وتكون جميع العينات الممكنة هى :

(١١،١١) ، (١٦،١١) ، (١٢،١١) ، (١٥،١١) ، (١٦،١١) ، (١٤،١١)
(١١،١٦) ، (١٦،١٦) ، (١٢،١٦) ، (١٥،١٦) ، (١٦،١٦) ، (١٤،١٦)
(١١،١٢) ، (١٦،١٢) ، (١٢،١٢) ، (١٥،١٢) ، (١٦،١٢) ، (١٤،١٢)
(١١،١٥) ، (١٦،١٥) ، (١٢،١٥) ، (١٥،١٥) ، (١٦،١٥) ، (١٤،١٥)
(١١،١٦) ، (١٦،١٦) ، (١٢،١٦) ، (١٥،١٦) ، (١٦،١٦) ، (١٤،١٦)
(١١،١٤) ، (١٦،١٤) ، (١٢،١٤) ، (١٥،١٤) ، (١٦،١٤) ، (١٤،١٤)

الأوساط الحسابية لهذه العينات هى :

١١ ، ١٣,٥ ، ١١,٥ ، ١٣ ، ١٣,٥ ، ١٢,٥
١٣,٥ ، ١٦ ، ١٤ ، ١٥,٥ ، ١٦ ، ١٥
١١,٥ ، ١٤ ، ١٢ ، ١٣,٥ ، ١٤ ، ١٣
١٣ ، ١٥,٥ ، ١٣,٥ ، ١٥ ، ١٥,٥ ، ١٤,٥
١٣,٥ ، ١٦ ، ١٤ ، ١٥,٥ ، ١٦ ، ١٥
١٢,٥ ، ١٥ ، ١٣ ، ١٤,٥ ، ١٥ ، ١٤

ويمكن تلخيص هذه الأوساط في التوزيع التكرارى المبين بالعمودين الأول والثانى من الجدول (٦ – ١) وهذا هو توزيع المعاينة المطلوب .

الجدول (٦ – ١)

توزيع المعاينة للوسط الحسابى لأوزان ٦ أرانب للعينات من الحجم ٢

$\bar{س}_ر$	$ك_ر$	خطأ التقدير $ك_ر(\bar{س}_ر - \mu)$
١١	١	– ٣
١١,٥	٢	– ٥
١٢	١	– ٢
١٢,٥	٢	– ٣
١٣	٤	– ٤
١٣,٥	٦	– ٣
١٤	٥	٠
١٤,٥	٢	١
١٥	٥	٥
١٥,٥	٤	٦
١٦	٤	٨
المجموع	٣٦	صفر

(ثانيا) الوسط الحسابى لتوزيع المعاينة = $\mu_{\bar{س}} = \frac{1}{ن} \sum ك_ر \bar{س}_ر$

أى أن $\mu_{\bar{س}} = \frac{1}{٣٦} (١١×١ + ١١,٥×٢ + ... + ١٦×٤)$

$= \frac{٥٠٤}{٢٦} = ١٤$ أوقية

$$\text{تباين توزيع المعاينة} = \frac{1}{\text{ن}} \left[\text{مج ك}_\text{ر}\ \bar{\text{س}}_\text{ر}^{2} - \frac{(\text{مج ك}_\text{ر}\ \bar{\text{س}}_\text{ر})^{2}}{\text{ن}} \right]$$

$$\text{أى أن } \sigma^{2}_{\bar{\text{س}}} = \frac{1}{36} \left[(1 \times 11^{2} + 2 \times 11{,}5^{2} + \ldots + 4 \times 16^{2}) - \frac{504^{2}}{36} \right]$$

$$= \frac{11}{6}$$

$$\text{الوسط الحسابى للمجتمع } \mu = \frac{1}{6}(11+16+12+15+16+14) = 14$$

$$\text{تباين المجتمع } \sigma^{2} = \frac{1}{6}\left[(11^{2} + 16^{2} + \ldots + 14^{2}) - \frac{84^{2}}{6}\right] = \frac{11}{3}$$

بالمقارنة نجد ما يلى :

(أ) $\mu_{\bar{\text{س}}} = \mu$

أى أن الوسط الحسابى لتوزيع المعاينة = الوسط الحسابى للمجتمع .

(ب) $\sigma^{2}_{\bar{\text{س}}} = \frac{\sigma^{2}}{\text{ن}}$

أى أن تباين توزيع المعاينة = تباين المجتمع مقسوما على حجم العينة .

وهاتان النتيجتان هما قاعدتان عامتان بالنسبة لتوزيع المعاينة للوسط الحسابى ويمكن إثباتهما رياضيا ، سواء كان المجتمع منتهيا والمعاينة مع الإرجاع أو كان لا نهائيا .

ملاحظة :

الأوساط الحسابية للعينات المأخوذة من مجتمع ما تختلف بطبيعة الحال عن بعضها البعض ، وحين نأخذ أحد هذه الأوساط $\bar{\text{س}}_\text{ر}$ لتقدير الوسط الحسابى μ للمجتمع يكون هناك خطأ فى التقدير قدره $\bar{\text{س}}_\text{ر} - \mu$ ففى هذا المثال لدينا μ = ١٤ وإذا قدرنا هذا المتوسط من متوسط العينة الأولى وهو ١١ يكون هناك خطأ قدره ١١−١٤ = −٣ وحين نقدره من متوسط العينة الثانية وهو ١١,٥

يكون هناك خطأ قدره ٢ (١١,٥ − ١٤) = ٢ × − ٢,٥ = −٥ (مع ملاحظة أن هناك عينتين متوسط كل منهما ١١,٥) وهكذا بالنسبة لأخطاء التقدير من متوسطات العينات الأخرى كما هو مبين بالعمود الثالث من الجدول (٦ − ١) ، حيث نلاحظ أيضا أن مجموع أخطاء التقدير هذه يساوى صفرا وهذا هو الذى أدى إلى المساواة (أ) .

(٦ − ٢ − ١) الإحصاءة A statistic

فى المثال السابق كان لدينا ٣٦ وسطا حسابيا $\bar{س}_١$ ، $\bar{س}_٢$ ، ..... ، $\bar{س}_{٣٦}$ وسمينا التوزيع التكرارى لهذه المتوسطات بتوزيع المعاينة للوسط الحسابى للعينات ذوات الحجم ٢. وحين نسحب عينة ما من مجتمع الأرانب فإننا لا نعلم مقدما المتوسط الذى نحصل عليه منها بل يكون ذلك متروكا للصدفة . ولذلك ننظر إلى هذه المتوسطات على أنها قيم مشاهدة من متغير عشوائى $\bar{س}$ ونسمى هذا المتغير حينئذ بالإحصاءة $\bar{س}$ ويكون التوزيع سابق الذكر هو توزيع المعاينة لهذه الإحصاءة .

كذلك إذا كانت $ع^٢_١$ ، $ع^٢_٢$ ، ... هى تباينات جميع العينات التى من نفس الحجم التى يمكن أن تؤخذ من مجتمع ما فإننا ننظر إليها كقيم متغير عشوائى $_مع^٢$ ويكون توزيع هذه القيم هو توزيع المعاينة للإحصاءة $_مع^٢$. وبالمثل لأى مقياس إحصائى آخر .

(٦ − ٢ − ٢) الخطأ المعيارى Standard Error

يسمى الانحراف المعيارى لتوزيع المعاينة لإحصاءة ما بالخطأ المعيارى لهذه الإحصاءة . ففى المثال السابق الخطأ المعيارى للوسط الحسابى أو للاحصاءة $\bar{س}$ هو $\sigma_{\bar{س}} = \sqrt{\frac{٢٠}{٦}} = ١,٨٣$ تقريبا . والأخطاء المعيارية تلعب دورا رئيسيا فى نظرية العينات كما سنتبين بعد .

(٦ – ٢ – ٣) التقدير غير المتحيز Unbiased Estimate

نظرا لأن الوسط الحسابى $\mu_{\bar{س}}$ للإحصاءة $\bar{س}$ يساوى الوسط الحسابى μ للمجتمع ، نقول إن $\bar{س}$ هو مُقدِّر غير متحيز للوسط الحسابى للمجتمع unbiased estimator كما نقول للوسط الحسابى $\bar{س}_ر$ لأى عينة عشوائية بأنه تقدير غير متحيز لمتوسط المجتمع ، وأى فرق $\bar{س}_ر - \mu$ يعتبر كما سبق القول خطأ فى تقدير متوسط المجتمع من متوسط العينة ، ويقاس مدى دقة هذا التقدير بالخطأ المعيارى $\sigma_{\bar{س}} = \frac{\sigma}{\sqrt{ن}}$ لتوزيع المعاينة للاحصاءة $\bar{س}$. لاحظ أن هذا الخطأ يتوقف على كل من σ ، ن ويكون هذا الخطأ صغيرا إذا كانت σ صغيرة أو كان حجم العينة ن كبيرا ومن ثم كان كبر العينة من العوامل الرئيسية لدقة التقدير .

وبصفة عامة إذا كنا نقدر أحد بارامترات مجتمع من عينة فإن التقدير يوصف بأنه غير متحيز إذا كان متوسط قيم هذا التقدير على جميع العينات العشوائية التى يمكن أخذها من المجتمع يساوى البارامتر الذى نقدره ، ومن الواضح أن هذه الصفة مطلوبة فى التقدير . وكما وجدنا أن $\bar{س}$ تقدير غير متحيز للبارامتر μ نجد أن $ع^2 = \frac{1}{ن - 1}$ مجـ $(س - \bar{س})^2$ تقدير غير متحيز للتباين σ^2 للمجتمع .

(٦ – ٣) المعاينة من مجتمعات معتدلة

Sampling From Normal Populations

إن الهدف الرئيسى من دراسة الإحصاء معرفة كيفية الحكم على المجتمعات وإصدار قرارات عنها عن طريق عينات مأخوذة منها ، وهذا ما سميناه بعملية الاستدلال الإحصائى . ولما كانت هذه العملية تتوقف على معرفتنا بتوزيعات الاحتمال للإحصاءات التى نستخدمها ، وجب علينا أن نحيط بهذه التوزيعات توطئة لتحقيق ذلك الهدف .

ومن أهم توزيعات المعاينة تلك التى تكون فيها المعاينة من مجتمعات معتدلة ، وسنقدم فى هذا البند عددا من هذه التوزيعات دون التعرض للبراهين الرياضية .

(أولا) توزيع المعاينة للوسط الحسابى

Sampling Distribution of the Mean

إذا كان لدينا مجتمع معتدل وسطه الحسابى μ وانحرافه المعيارى σ فإن توزيع المعاينة للإحصاءة $\bar{س}$ للعينات ذوات الحجم ن المأخوذة من هذا المجتمع يكون توزيعا **معتدلا** وسطه الحسابى μ وانحرافه المعيارى $\frac{\sigma}{\sqrt{ن}}$ (أى الخطأ المعيارى) ، وبالتالى فإن الإحصاءة

$$ص_1 = \frac{\bar{س} - \mu}{\sigma / \sqrt{ن}} \qquad (1)$$

يكون توزيعها مطابقاً للتوزيع المعتدل المعيارى . راجع الخاصة (ذ) بالبند (٤ – ١) . والمفروض في هذه الإحصاءة أن تكون كل من μ ، σ معروفة القيمة .

كما أن الإحصاءة

$$ص_2 = \frac{\bar{س} - \mu}{ع_ن / \sqrt{ن}} \qquad (2)$$

يكون توزيعها مطابقاً لتوزيع ت بدرجات حرية عددها ن – ١ . وتستخدم هذه الإحصاءة حينما تكون σ^2 مجهولة القيمة وهذا ما يحدث في أغلب الحالات ، ونضطر حينئذ لاستخدام تباين العينة وهو

$$ع^2 = \frac{1}{ن-1} \left(مجـ س^2 - \frac{(مجـ س)^2}{ن} \right)$$ كتقدير غير متحيز للتباين σ^2 للمجتمع .

(ثانياً) توزيع المعاينة للفرق بين وسطين حسابيين :

نفرض أن لدينا مجتمعين معتدلين وسطاهما الحسابيان $\mu_{١}$ ، $\mu_{٢}$ وانحرافاهما المعياريان $\sigma_{١}$ ، $\sigma_{٢}$ ونفرض أن الإحصاءة $\bar{س}_{١}$ ترمز إلى الوسط الحسابي لعينة عشوائية حجمها $ن_{١}$ مأخوذة من المجتمع الأول وأن الإحصاءة $\bar{س}_{٢}$ ترمز إلى الوسط الحسابي لعينة عشوائية حجمها $ن_{٢}$ مأخوذة من المجتمع الثاني . إذا كانت هاتان العينتان مستقلتين فإن توزيع المعاينة للإحصاءة $(\bar{س}_{١}-\bar{س}_{٢})$ يكون توزيعاً معتدلا وسطه الحسابي $(\mu_{١}-\mu_{٢})$ وانحرافه المعياري $\sqrt{\frac{\sigma_{١}^{٢}}{ن_{١}}+\frac{\sigma_{٢}^{٢}}{ن_{٢}}}$ وبالتالى فإن الإحصاءة :

$$ص_{٣} = \frac{(\bar{س}_{١}-\bar{س}_{٢})-(\mu_{١}-\mu_{٢})}{\sqrt{\frac{\sigma_{١}^{٢}}{ن_{١}}+\frac{\sigma_{٢}^{٢}}{ن_{٢}}}} \qquad (٣)$$

يكون توزيعها مطابقاً للتوزيع المعتدل المعيارى .

والمفروض هنا أن تكون كل من $\mu_{١}$ ، $\mu_{٢}$ ، $\sigma_{١}$ ، $\sigma_{٢}$ معروفة القيمة .

كما أن الإحصاءة

$$ص_{٤} = \frac{(\bar{س}_{١}-\bar{س}_{٢})-(\mu_{١}-\mu_{٢})}{ع\sqrt{\frac{١}{ن_{١}}+\frac{١}{ن_{٢}}}} \qquad (٤)$$

يكون توزيعها مطابقاً لتوزيع ت بدرجات حرية عددها $ن_{١}+ن_{٢}-٢$

$$\text{حيث } ع_م^2 = \frac{(ن_1 - 1) ع_{1}^2 + (ن_2 - 1) ع_{2}^2}{ن_1 + ن_2 - 2} \qquad (أ)$$

$$= \frac{مج(س_{1م} - \bar{س}_1)^2 + مج(س_{2م} - \bar{س}_2)^2}{ن_1 + ن_2 - 2} \qquad (أَ)$$

وحيث $\sigma_1 = \sigma_2$

أما إذا كان $\sigma_1 \neq \sigma_2$ فنستخدم الاحصاءة

$$ص_\circ = \frac{(\bar{س}_1 - \bar{س}_2) - (\mu_1 - \mu_2)}{\sqrt{ع_1^2/ن_1 + ع_2^2/ن_2}} \qquad (٥)$$

التى لها توزيع ت بدرجات حرية $\left(\frac{ع_1^2}{ن_1} + \frac{ع_2^2}{ن_2}\right)^2 \div$

$$\left[\left(\frac{ع_1^2}{ن_1}\right)^2 \div (ن_1 - 1) + \left(\frac{ع_2^2}{ن_2}\right)^2 \div (ن_2 - 1)\right]$$

ملاحظة :

في المعاينة من أى مجتمع (ليس معتدلا بالضرورة) نستطيع رياضياً إثبات ما يلى اعتماداً على النظرية الشهيرة المسماه بنظرية النهاية المركزية central limit theorem .

(أ) يقترب توزيع الإحصاءة $ص_1$ المعرفة في (١) من التوزيع المعتدل المعيارى حين يقترب حجم العينة من اللانهاية .

(ب) يقترب توزيع الإحصاءة $ص_٣$ المعرفة في (٣) من التوزيع المعتدل المعيارى حين تقترب كل من $ن_١$ ، $ن_٢$ من اللانهاية .

إن المعني التطبيقي لهاتين النظريتين أنه عند المعاينة من مجتمع غير معتدل وبشرط أن تكون العينات كبيرة الحجم (أكبر من ٣٠) يجوز عملياً اعتبار أن توزيع كل من $ص_١$ ، $ص_٣$ هو بالتقريب توزيع معتدل معيارى ، وكلما كبر حجم العينات كلما قل الخطأ الناشيء عن هذا التقريب .

(ثالثاً) توزيع المعاينة لخارج قسمة تباينين :

إذا كانت الإحصاءتان $ع^٢_١$ ، $ع^٢_٢$ ترمزان إلى تبايني عينتين عشوائيتين مستقلتين مأخوذتين من مجتمعين معتدلين لهما نفس التباين ($\sigma^٢_١ = \sigma^٢_٢$) فإن الإحصاءة

$$ص_٦ = \frac{ع^٢_١}{ع^٢_٢} \qquad (٦)$$

يكون توزيعها مطابقاً لتوزيع ف بدرجتي حرية $ن_١ - ١$ ، $ن_٢ - ١$ حيث $ن_١$ ، $ن_٢$ هما حجما العينتين. ويلاحظ أنه إذا كانت $ع^٢_١$ ، $ع^٢_٢$ قيمتين مأخوذتين من عينتين عشوائيتين مستقلتين من مجتمع واحد أو من مجتمعين لهما نفس التباين فإن النسبة $ع^٢_١ / ع^٢_٢$ تكون قريبة من الواحد لأن كلا من $ع^٢_١$ ، $ع^٢_٢$ تقدير لنفس التباين $\sigma^٢$.

(٦ - ٤) المعاينة من توزيع ذى الحدين :

SAMPLING FROM A BINOMIAL DISTRIBUTION

اعتبر مجتمعاً مقسماً إلى قسمين منفصلين من حيث وقوع أو عدم وقوع حدث معين أ . افرض أن احتمال وقوع هذا الحدث في أى تجربة واحدة هو مقدار ثابت ح واحتمال عدم وقوعه ك = ١ - ح . حسب البند (٣ - ٣) وتحت الشروط

المذكورة يكون توزيع عدد مرات وقوع الحدث في $ن$ من التجارب هو توزيع ذى الحدين دليلاه $ن$ ، $ح$،وسطه الحسابى $ن ح$ وانحرافه المعيارى $\sqrt{ن ح ك}$ ويكون التكرار النسبي لوقوع هذا الحدث هو خارج قسمة عدد مرات وقوع الحدث على العدد $ن$ وعلى ذلك فإن متوسط هذه النسبة يساوى :

$$\frac{ن ح}{ن} = ح \text{ وانحرافها المعيارى } \frac{\sqrt{ن ح ك}}{ن} = \sqrt{\frac{ح ك}{ن}}$$

وإذا أخذنا من هذا المجتمع جميع العينات ذوات الحجم $ن$ وحسبنا في كل منها القيمة $س$ لعدد مرات وقوع ذلك الحدث ثم حسبنا نسبة هذا العدد وهو $ر = \frac{س}{ن}$ فإننا نحصل على توزيع المعاينة لهذه النسبة .

وقد رأينا فى البند (٤ — ٦) أنه إذا كانت $ن$ كبيرة ولم تكن أى من $ح$ أو $ك$ قريبة من الصفر فإن :

(أ) توزيع الاحتمال لعدد مرات وقوع الحدث في العينات ذوات الحجم $ن$ يقترب من توزيع معتدل وسطه الحسابي $ن ح$ ، وانحرافه المعيارى $\sqrt{ن ح ك}$ وبالتالى فإن توزيع الإحصاءة

$$ص_٧ = \frac{س - ن ح}{\sqrt{ن ح ك}} \qquad (٧)$$

يقترب من التوزيع المعتدل المعيارى مع (٠ ، ١) .

(ب) وبالمثل فإن توزيع الإحصاءة :

$$ص_٨ = \frac{ر - ح}{\sqrt{\frac{ح ك}{ن}}} \qquad (٨)$$

يقترب من التوزيع المعتدل المعيارى .

(٦ – ٥) اختبارات الفروض : TESTS OF HYPOTHESES

الفرض الإحصائي : هو جملة أو مقولة نذكرها عن مجتمع أو عدة مجتمعات بهدف اختبار صواب أو خطأ هذه الجملة ، ولمعرفة هذا الصواب أو هذا الخطأ نستعين بما يسمى باختبارات الفروض وهى اختبارات تبني على إحصاءات تكون توزيعاتها معروفة لنا مثل تلك التي وردت في البندين (٦ – ٣) و(٦ – ٤) السابقين .

واختبار الفرض هو بكل بساطة قاعدة تؤدى إلى اتخاذ قرار برفض أو قبول الفرض فور حصولنا على قيم مشاهدة في عينة عشوائية . وفي المعتاد نضع فرضاً نرمز له بالرمز ف٠ . نسميه **بالفرض الصفرى** أو بفرض العدم null hypothesis نصوغه بحيث يعبر عن انعدام الفرق بين شيئين ، مثلا $\mu_1 = \mu_2$ (أى $\mu_1 - \mu_2 = 0$) أو $\mu =$ أ (أى $\mu -$ أ $= 0$) أو لا توجد علاقة بين متغيرين . وأى فرض يراد اختباره ضد الفرض الصفرى يسمى بالفرض الآخر أو بالفرض البديل alternative hypothesis ويرمز له بالرمز ف١ وقد يأخذ هذا الفرض الشكل . $\mu_1 \neq \mu_2$ أو $\mu_1 > \mu_2$ أو $\mu_1 < \mu_2$.

ولاختبار الفرض الصفرى ف٠ ضد الفرض الآخر ف١ نبدأ باختيار إحصاءة تناسب الفرض المختبر بشرط أن يكون توزيع المعاينة لهذه الإحصاءة معلوماً لنا ثم نحدد في هذا التوزيع منطقة م يكون احتمال وقوع القيم المشاهدة لهذه الإحصاءة فيها هو احتمال صغير α ، مثلا ٠,٠٥ ، ٠,٠١ ، ٠,٠٠١ ، ... وتسمى هذه المنطقة بالمنطقة الحرجة أو بمنطقة الرفض critical region or region of rejection أما الاحتمال α فيسمى بمستوى الدلالة للاختبار level of significance وتأخذ القاعدة التي يعطيها الاختبار الصورة الآتية :

« إذا وقعت قيمة مشاهدة لهذه الإحصاءة ، محسوبة على أساس صحة الفرض الصفرى ف٠ ، داخل المنطقة الحرجة نرفض ف٠ عند مستوى الدلالة α وإلا نقبل ف٠ » .

وذلك على أساس أن احتمال وقوع القيمة المشاهدة داخل المنطقة الحرجة هو احتمال ضئيل يجعلنا نشك في صحة الفرض الصفرى . ونظراً لأن القيم المشاهدة يمكن أن تقع في أى جزء من التوزيع فإن هذه القاعدة تعنى أننا نغامر برفض الفرض الصفرى (عند المستوى α) مع علمنا بأننا قد نكون مخطئين في رفضنا هذا ، وإن كان احتمال هذا الخطأ هو احتمال صغير لا يزيد عن α . أما إذا وقعت القيمة المشاهدة في المنطقة المكملة للمنطقة الحرجة ، وتسمى بمنطقة القبول ، فلا يسعنا إلا قبول الفرض $ف_٠$ على أساس أن احتمال وقوع القيمة المشاهدة فيها هو احتمال كبير ١ − α .

وجدير بالملاحظة أنه بينما يمكننا الاختبار من هدم الفرض الصفرى إلا أنه لا يستطيع أن يثبت صحته وكل ما يستطيع أن يفعله لصالح هذا الفرض هو أن يبين عدم وجود ما يتعارض معه من واقع العينة المستخدمة ، وحين نقول إننا نقبل $ف_٠$ لا نعنى أننا أثبتنا صحته وإنما نعنى أن ما لدينا من بيانات لا تعطى دليلا كافيا يدعو إلى رفضه ، وحين نقول إننا نرفض $ف_٠$ فإننا لا نعنى أننا أثبتنا زيفه وإنما نعنى أن ما لدينا من بيانات لا تدعم صحته بل تشير إلى صحة الفرض الآخر $ف_١$.

واختبار الفرض $ف_٠$ عن مجتمع ضد فرض آخر $ف_١$ يستلزم بطبيعة الحال الحصول على عينة عشوائية من هذا المجتمع ، ويفضل اختيار قيمة α حتى قبل اختيار العينة . ويتوقف تحديدنا لقيمة α على طبيعة المشكلة التي نتناولها ودرجة المغامرة التي نقبلها لتحمل مسئولية الخطأ المحتمل . وبصفة عامة يمكن تلخيص خطوات إجراء اختبار الفروض فيما يلى :

(أولا) نحدد الفرض الصفرى $ف_٠$ والفرض الآخر $ف_١$ من واقع المشكلة التي نتناولها .

(ثانياً) نحدد الإحصاءة التى تناسب الفرض المختبر ونحدد المنطقة الحرجة م فى توزيع هذه الإحصاءة بناء على مستوى الدلالة α السابق اختياره .

(ثالثاً) نحسب قيمة الإحصاءة من واقع البيانات المشاهدة في عينة عشوائية وعلى أساس أن الفرض الصفرى صحيح .

(رابعاً) الاستنتاج : نتخذ قراراً برفض أو قبول الفرض الصفرى ف عند مستوى الدلالة α بحسب وقوع القيمة المحسوبة للإحصاءة داخل أو خارج المنطقة الحرجة .

ويجدر الإشارة هنا إلى أن اختبارات الفروض ما هى إلا وسيلة حسابية تستخدم البيانات التى حصلنا عليها من العينات لإلقاء الضوء على صحة أو خطأ الفرض الصفرى ، وينبغى للباحث عند إصدار قراره أن يضيف إلى نتيجة الاختبار جميع ما لديه من معلومات وخبرات وأبحاث سابقة عن موضوع الدراسة .

كما يجدر الإشارة إلى أن حجم العينة يلعب دورا هاما فى عملية الاستدلال الاحصائى . وحين تكون العينة صغيرة فإن اختبار الدلالة لا يؤدى إلى رفض الفرض الصفرى إلا إذا كان هذا الفرض خاطئا بدرجة كبيرة ، وبالعكس حين تكون العينة كبيرة فإن أى انحراف صغير عن الفرض الصفرى يسهل اكتشافه كانحراف ذى دلالة . وبصفة عامة يفضل أن يكون حجم العينة كبيرا بدرجة كافية منعا للوقوع فيما يسمى بالخطأ من النوع الثانى ، وهذا ما سوف نتناوله فى الفصل السابع .

في البنود (٦ – ٦) ، (٦ – ٧) (٦ – ٨) الآتية نتناول ثلاثة من أشهر اختبارات الفروض تعرف باختبار ت واختبار χ^2 (كا٢) واختبار ف .

(٦ – ٦) اختبار ت : THE t-TEST

هناك عدة إحصاءات تطابق توزيعاتها توزيع ت السابق دراسته . وإذا بني اختبار فروض على أى من هذه الإحصاءات قيل إنه اختبار ت . والصورة العامة لهذه الإحصاءات هى :

$$ت = \frac{\text{القيمة المشاهدة للاحصاءة – الوسط الحسابى للإحصاءة}}{\text{تقدير للخطأ المعيارى للاحصاءة}}$$

بشرط أن يكون للإحصاءة توزيع معتدل وأن تكون القيمة المشاهدة هي تقدير غير متحيز لمتوسط الإحصاءة .

(٦ – ٦ – ١) اختبار فرض عن الوسط الحسابي لمجتمع معتدل :

مثال (٦ – ٢) :

تقضي التعليمات الحكومية بأن تكون الجرعة القياسية من مستحضر بيولوجي ٦٠٠ وحدة نشاط في السنتيمتر المكعب . اختبرت عينة عشوائية حجمها ١٠ من إنتاج شركة ما ووجد أن متوسطها ٥٩٢,٥ وحدة نشاط في السنتيمتر المكعب بانحراف معيارى ١١,٢ وحدة . هل نستطيع القول بأن إنتاج هذه الشركة يتمشي مع التعليمات الحكومية ؟

الحل :

نريد أن نختبر ما إذا كان الوسط الحسابي μ لمجتمع المستحضر الذى تنتجه الشركة يساوى الجرعة القياسية التي حددتها التعليمات الحكومية وهى ٦٠٠ وحدة نشاط / سم^3. نتبع الخطوات الأربع المشار إليها فى البند السابق .

(١) الفرض الصفرى ف$_\cdot$: μ = ٦٠٠

الفرض الآخر ف$_1$: $\mu \neq$ ٦٠٠

، α = ٠,٠٥

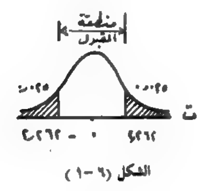
(٢) لنهتدى إلى الإحصاءة المناسبة نذكر أننا نريد اختبار فرض عن قيمة الوسط الحسابي μ للمجتمع عن طريق قيمة الوسط الحسابي $\bar{س}$ لعينة ونتوقع أنه إذا كان هذا الفرض صحيحاً فإن $\bar{س}$ تكون قريبة من μ ، أى أننا نريد اختبار ما إذا كان الفرق بين μ ، $\bar{س}$ هو فرق صغير نعزوه إلى الصدفة أو فرق جوهرى (ذو دلالة)

يدعونا إلى عدم الثقة في القيمة المفروضة للدليل μ . (هذا على أساس أن العينة هي عينة عشوائية ممثلة للمجتمع تمثيلا جيداً وبالتالى فإن وسطها الحسابي $\bar{س}$ هو عدد نثق فيه) . هذا التحليل يشير مباشرة إلى أن أنسب إحصاءة لقياس صغر أو كبر هذا الفرق هى الإحصاءة (٢) وهى :

$$ت = \frac{\bar{س} - \mu}{ع/\sqrt{ن}}$$

التي نعلم أن توزيعها يطابق توزيع ت بدرجات حرية عددها $\nu = ن - ١$ ، مع ملاحظة أن هذه الإحصاءة تنعدم عندما تتساوى قيمتي μ ، $\bar{س}$ وتكبر قيمتها المطلقة كلما كبر الفرق بينهما .

المنطقة الحرجة م = $\{ت : |ت| > ت_{\alpha(ن - ١)}\}$



الشكل (٦-١)

وهى تتألف من جزءين واقعين أسفل جانبي منحني ت واحتمال كل منهما $\frac{\alpha}{٢}$ وبالتالى فإن احتمال وقوع قيمة الإحصاءة في هذه المنطقة هو احتمال صغير α .

(٣) نحسب قيمة هذه الإحصاءة من بيانات العينة وعلى أساس صحة الفرض الصفرى . وإذا رمزنا لهذه القيمة بالرمز $ت_ى$ نجد أن :

$$ت_ى = \frac{٥٩٢٫٥ - ٦٠٠}{١١٫٢ / \sqrt{١٠}} = -٢٫١٢ \quad \text{حيث} \quad \nu = ١٠ - ١ = ٩$$

فإذا كنا قد اخترنا α = ٠,٠٥ نجد من جدول ت أن القيمة الحرجة التى تحدد منطقة الرفض هى : $ت_{٠,٠٠٥ [٩]}$ = ٢,٢٦٢

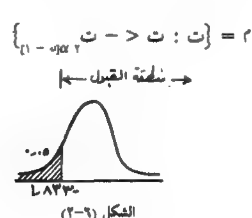
(٤) الاستنتاج : بما أن $ت_{ى}$ = −٢,١٢ لا تقع في المنطقة الحرجة لا يكون لدينا دليل ضد الفرض الصفرى عند مستوى الدلالة ٠,٠٥ أى أننا نقبل الفرض أن μ = ٦٠٠ ونقرر بأن إنتاج الشركة يتمشى مع التعليمات الحكومية .

الاختبار ذو الجانب الواحد وذو الجانبين :

في المثال السابق اختبرنا الفرض الصفرى $ف_{٠}$: μ = ٦٠٠ ضد الفرض $ف_{١}$: $\mu \neq$ ٦٠٠ وعلى ذلك كان لنا أن نرفض $ف_{٠}$ في حالتين هما : أن تكون μ < ٦٠٠ أو تكون μ > ٦٠٠ ولذلك شملت المنطقة الحرجة كلا جانبي توزيع ت . ونقول حينئذ إن الاختبار ذو جانبين أو إنه اختبار غير اتجاهى .

ولكن إذا كانت زيادة وحدات النشاط في المستحضر البيولوجي عن المستوى القياسي لا ينتج عنها ضرر بينما النقصان عنه يفقد المستحضر فعاليته فإن اهتمامنا يكون منصباً على ما إذا كان متوسط العينة أصغر صغراً ذا دلالة من المستوى القياسي . في هذه الحالة يكون المطلوب اختبار الفرض الصفرى $ف_{٠}$: μ = ٦٠٠ ضد الفرض $ف_{١}$: μ < ٦٠٠ وتكون المنطقة الحرجة فى جانب واحد فقط من التوزيع هو الجانب الأيسر ، أى أن المنطقة الحرجة تكون في هذه الحالة على الصورة ،

م = {ت : ت < − $ت_{٢\alpha [ن - ١]}$}



الشكل (٦-٢)

ونقول حينئذ إن الاختبار ذو جانب واحد أو إنه اختبار اتجاهي . فإذا كان مستوى الدلالة هو α = ٠,٠٥ فإن القيمة التي تحد المنطقة الحرجة من اليمين وتجعل المساحة بينها وبين المنحنى ٠,٠٥ يمكن إيجادها من جدول ت باستخدام الاحتمال $\acute{\alpha}$ = ٢ α = ٠,١٠ .

من الجدول نجد أن $ت_{٠,١ [٩]}$ = ١,٨٣٣

إذن القيمة التي تحد المنطقة الحرجة من اليمين هى – ١,٨٣٣.

بما أن –٢,١٢ < –١,٨٣٣ فإن القيمة المشاهدة –٢,١٢ تقع في المنطقة الحرجة وبالتالى نرفض الفرض الصفرى ف. عند مستوى الدلالة ٠,٠٥ لصالح الفرض الآخر ونقرر أن متوسط إنتاج الشركة يقل عن المستوى الذى تحدده التعليمات الحكومية .

ولا نعجب من اختلاف هذه النتيجة عن النتيجة السابقة لأن كلا منهما يجيب على سؤال مختلف هو السؤال الذى تتطلب المشكلة الإجابة عنه .

وبالمثل يمكن أن يكون اهتمامنا منصباً على الجانب الأيمن فقط من التوزيع حيث تكون المنطقة الحرجة على الصورة م = {ت : ت> $ت_{٢\alpha(ن-١)}$} . وفي تناول أى مشكلة علينا أن نفكر جيداً قبل أن نحدد ما إذا كانت تتطلب اختباراً ذا جانب واحد أو ذا جانبين تحسباً من الوقوع في خطأ في عملية الاستدلال . وهذا التحديد ينبغى أن يتقرر عند تصميم التجربة وقبل جمع البيانات وبحسب التساؤل الذى تطرحه المشكلة . فمثلا إذا كنا نقارن أداء مجموعة من الطلاب بمستوى معروف فإن اهتمامنا ينصب على معرفة ما إذا كانت المجموعة أحسن أو أسوأ من هذا المستوى ، وهنا يجب أن يكون الاختبار ذا جانبين . أما إذا كنا نقارن نوعا جديدا من الدواء بنوع تقليدى فإن اهتمامنا يكون منصبا على معرفة ما إذا كان الدواء الجديد أفضل من التقليدى وهنا يجب أن يكون الاختبار ذا جانب واحد .

ملاحظة (١) :

يتطلب استخدام اختبار ت بهذه الصورة تحقق الشرطين الآتيين :

(أ) أن تكون العينة عشوائية بسيطة .

(ب) أن يكون المجتمع معتدلا (أو يمكن اعتباره معتدلا) .

(٦ - ٢ - ٢) فترات الثقة للوسط الحسابي لمجتمع معتدل :

CONFIDENCE INTERVALS

إذا كان الوسط الحسابي μ لمجتمع ما مجهولا ، يمكن تقديره بواسطة الوسط الحسابي $\bar{س}$ لعينة عشوائية ذات حجم مناسب مأخوذة من هذا المجتمع ، ومثل هذا التقدير يسمى **بالتقدير بنقطة** point estimation ، ولكن نظراً لأن متوسطات العينات المأخوذة من نفس المجتمع تختلف من عينة إلى أخرى حتي ولو كانت العينات من نفس الحجم فإن هذا التقدير يشوبه بعض الصعوبات خاصة في تحديد مدى الثقة التي نضعها فيه .

ولذلك يفضل في كثير من الأحيان تقدير الوسط الحسابي وغيره من أدلة المجتمع عن طريق ما يسمى **بالتقدير بفترة** interval estimation حيث تحدد فترة يقع فيها الدليل المجهول بدرجة ثقة معينة .

حينما يكون المجتمع **معتدلا** وسحبنا منه عينة عشوائية بسيطة حجمها ن ووسطها الحسابي $\bar{س}$ وانحرافها المعياري ع فإن الفترة :

$$\left(\bar{س} - \frac{ع}{\sqrt{ن}}\, ت_{\alpha[ن-1]} \text{ ، } \bar{س} + \frac{ع}{\sqrt{ن}}\, ت_{\alpha[ن-1]}\right) \qquad (١٠)$$

تسمى بفترة ثقة بدرجة $(1 - \alpha)$ لمتوسط المجتمع μ . وهذه العبارة لا تعني أن احتمال وقوع μ في هذه الفترة هو $(1 - \alpha)$ لأن μ عدد ثابت لا توزيع له وإنما تفسر هذه العبارة كما يلي :

« إذا سحبنا جميع العينات التي من نفس الحجم وأوجدنا فترة الثقة بدرجة $(1 - \alpha)$ لكل منها فإن $(1 - \alpha)$ من هذه الفترات تشمل البارامتر μ » ، وهذا ما يمكن إثباته رياضياً .

ويسمى العددان $\bar{س} \pm \frac{ع}{\sqrt{ن}} ت_{\alpha[ن - 1]}$ اللذان يحدان فترة الثقة بحدى الثقة

بدرجة $(1 - \alpha)$ أما العدد $\frac{ع}{\sqrt{ن}}$ فهو تقدير للخطأ المعياري للإحصاءة $\bar{س}$ ،

كما تسمى α بمعامل الثقة .

مثال (٦ - ٣) :

أوجد فترة ثقة بدرجة ٩٥٪ للوسط الحسابي μ لمجتمع المستحضر البيولوجي المذكور في المثال (٦ - ٢) من واقع البيانات المعطاة في هذا المثال .

الحل :

لدينا $ن = ١٠$ ، $\bar{س} = ٥٩٢,٥$ ، $ع = ١١,٢$ ، $١ - \alpha = ٠,٩٥$ ومنها $\alpha = ٠,٠٥$

$$ت_{٠,٠٥[٩]} = ٢,٢٦٢ \quad ، \quad \frac{ع}{\sqrt{ن}} = \frac{١١,٢}{\sqrt{١٠}} = ٣,٥٤$$

بالتعويض في (١٠) نجد أن :

الحد الأدني للفترة $= ٥٩٢,٥ - ٣,٥٤ \times ٢,٢٦٢ = ٥٨٤,٤٩$

الحد الأعلى للفترة = ٥٩٢,٥ + ٣,٥٤ × ٢,٢٦٢ = ٦٠٠,٥١

إذن الفترة (٥٨٤,٤٩ ، ٦٠٠,٥١) هى فترة ثقة بدرجة ٩٥٪ لمتوسط المجتمع .

ملاحظة (٢) :

في المثال (٦ – ٢) قبلنا الفرض أن μ = ٦٠٠ ضد الفرض $\mu \neq$ ٦٠٠ عند مستوى الدلالة ٠,٠٥ . ونلاحظ أن العدد ٦٠٠ يقع في فترة الثقة التي أوجدناها في المثال (٦ – ٣) ، والواقع أن أى عدد نفرضه عن متوسط المجتمع μ يكون مقبولاً عند المستوى ٠,٠٥ طالما كان واقعاً في هذه الفترة . تحقق من ذلك باختيار بعض القيم المناسبة واختبارها كما بالبند (٦ – ٦ – ١) .

(٦ – ٦ – ٣) مقارنة متوسطى مجتمعين معتدلين :

(أو اختبار دلالة الفرق بين متوسطى عينتين مستقلتين مأخوذتين من مجتمعين معتدلين) .

مثال (٦ – ٤) :

في دراسة عن نبات حنك السبع كان المطلوب مقارنة الوسط الحسابي للنباتات ذات الزهور الحمراء من النوع Suttons Eclipse بالوسط الحسابي للنباتات ذات الزهور البيضاء من النوع Suttons Internediate White وقد نتجت البيانات الآتية من عينتين عشوائيتين مستقلتين .

ذات الزهور الحمراء : $ن_١$ = ٤٣ ، $\bar{س}_١$ = ١٣٤,٧٦ ، $ع_١$ = ٣٢,٣٥

ذات الزهور البيضاء : $ن_٢$ = ٢١ ، $\bar{س}_٢$ = ١٦٩,٥٠ ، $ع_٢$ = ٦٢,٩٣

اختبر ما إذا كان الفرق المشاهد بين متوسطى العينتين يرجع إلى تقلبات العينات أو إلى وجود فرق حقيقي بين المتوسطين $\mu_١$ ، $\mu_٢$ للمجتمعين اللذين سحبت منهما العينتان . افترض اعتدالية المجتمعين وتساوى تباينهما وخذ α = ٠,٠٥

الحل :

الفرض الصفرى ف. : $\mu_1 = \mu_2$ (لا يوجد فرق بين متوسطى المجتمعين)

الفرض الآخر ف١ : $\mu_1 \neq \mu_2$ (اختبار ذو جانبين)

الإحصاءة المناسبة هنا هى الإحصاءة (٤) وهى :

$$ت = \frac{(\bar{س}_1 - \bar{س}_2) - (\mu_1 - \mu_2)}{ع_م \sqrt{\frac{1}{ن_1} + \frac{1}{ن_2}}}$$

التي نعلم أن توزيعها يطابق توزيع ت بدرجات حرية عددها $\nu = ن_1 + ن_2 - 2$

وعلماً بأن $ع_م^2 = \frac{(ن_1 - 1)\, ع_1^2 + (ن_2 - 1)\, ع_2^2}{ن_1 + ن_2 - 2}$

على فرض أن تبايني المجتمعين متساويان

المنطقة الحرجة م = $\{ت : |ت| > ت_{\alpha[ن_1 + ن_2 - 2]}\}$

نحسب ت ى من بيانات العينة وعلى أساس صحة الفرض الصفرى .

$$ع^2 = \frac{٤٢ \times ٣٢{,}٣٥^2 + ٢٠ \times ٦٢{,}٩٣^2}{٦٢} = ١٩٨٦{,}٤١٣٥$$

$$= ١٩٨٦{,}٤١٣٥$$

$$\therefore ع = ٤٤{,}٥٦٩$$

تقدير الخطأ المعيارى $= ع \sqrt{\frac{1}{ن_1} + \frac{1}{ن_2}} = ٤٤{,}٥٦٩ \sqrt{\frac{1}{٤٣} + \frac{1}{٢١}} = ١١{,}٨٥٥$

$\therefore ت_ى = \frac{١٣٤{,}٧٦ - ١٦٩{,}٥٠}{١١{,}٨٥٥} = -٢{,}٩٣$ حيث $\nu = ٤٣ + ٢١ - ٢ = ٦٢$

من جدول ت نجد أن $ت_{٠{,}٠٥[٦٢]} = ٢{,}٠٠٠$

بما أن −٢,٩٣ > −٢ فإن القيمة $ت_ى$ = −٢,٩٣ تقع في المنطقة الحرجة (في الجانب الأيسر) وإذن نرفض الفرض الصفرى $ف_٠$: $\mu_1 = \mu_2$ عند مستوى الدلالة ٠,٠٥ ونقرر أن العينتين مأخوذتان من مجتمعين مختلفين في وسطهما الحسابي .

ملاحظة (٣)

يتطلب استخدام اختبار ت بهذه الصورة تحقق الشروط الآتية :

(أ) أن تكون كل من العينتين عشوائية بسيطة .
(ب) أن تكون كل من العينتين مأخوذة من مجتمع معتدل .
(جـ) أن تكون العينتان مستقلتين .
(د) أن يكون تباينا المجتمعين متساويان .

إذا لم تكن العينتان مستقلتين فإن اختبار ت للمقارنة بين متوسطى المجتمعين يأخذ الصورة التي سترد في البند (٨ − ٧ − ١) القادم .

فترات الثقة للفرق بين متوسطى مجتمعين معتدلين :

يمكن إثبات أنه تحت الشروط المذكورة ، يصلح العددان :

$$(\bar{\text{س}}_1 - \bar{\text{س}}_2) \pm \text{خ.م.ت}_{\alpha[\text{ن}_1 + \text{ن}_2 - 2]} \qquad (11)$$

كحدين لفترة ثقة بدرجة $(1 - \alpha)$ للفرق $(\mu_1 - \mu_2)$ بين متوسطى المجتمعين، حيث $\text{خ.م} = \text{ع} \sqrt{\frac{1}{\text{ن}_1} + \frac{1}{\text{ن}_2}}$ هو تقدير للخطأ المعيارى للإحصاءة $\bar{\text{س}}_1 - \bar{\text{س}}_2$.

ففي المثال (٦ – ٤) نجد أن :

الحد الأدني لفترة الثقة = (١٦٩,٥ – ١٣٤,٧٦) – ١١,٨٥٥ × ٢ = ١١,٠٣

الحد الأعلى لفترة الثقة = (١٦٩,٥ – ١٣٤,٧٦) + ١١,٨٥٥ × ٢ = ٥٨,٤٥

أى أننا يمكن أن نأخذ الفترة (١١,٠٣ ، ٥٨,٤٥) كفترة ثقة بدرجة ٩٥٪ للفرق بين متوسطى المجتمعين .

تمارين (٦ – ١)

في المسائل الآتية اعتبر أن المجتمعات معتدلة وأن العينات عشوائية ومستقلة وأن مستوى الدلالة α = ٠,٠٥ ما لم يذكر غير ذلك .

(١) وزنت ١٩ عذراء جرادة فكانت الأوزان كما يلي بالجرامات :

٠,٦٧ ٠,٩٥ ١,١٨ ١,٠٥ ١,٢٢ ٠,٨٢ ٠,٩٣ ٠,٦٨ ٠,٥٢ ٠,٧٠

٠,٥٧ ٠,٥٩ ٠,٣٥ ٠,٩٢ ٠,٦٧ ٠,٥٤ ١,١٢ ٠,٨٩ ٠,٨٤

(أ) اختبر الفرض أن الوسط الحسابي للوزن في مجتمع هذه الحشرة هو μ = ١ ضد كل من الفرضين الآتيين (١) $\mu \neq$ ١ ، (٢) $\mu <$ ١

(ب) أوجد فترة ثقة بدرجة ٩٥٪ للدليل μ .

(٢) باستخدام عينة حجمها ن = ١٥ ووسطها الحسابي $\overline{س}$ = ٨٥,٣ وتباينها ع^{٢} = ١٣ اختبر الفرض أن متوسط المجتمع μ = ٨٠ وأوجد فترة ثقة بدرجة ٩٩٪ لهذا المتوسط .

(٣) في عينة عشوائية من ٢٠ شخصاً يعالجون من مرض معين وجد أن عدد الكرات الدموية البيضاء مقدرة بالآلاف كالآتي . أوجد فترة ثقة بدرجة ٩٥٪ لمتوسط عدد الكرات البيضاء في مجتمع هؤلاء المرضي .

٨ ١٣ ٧ ٨ ١٣ ٦ ٤ ١٢ ٧ ٩ ١٠ ٧ ٨

١٠ ١٢ ١٣ ١٥ ١٠ ٥ ١١

(٤) في بحث ثبات reliability إحدى طرق القياس في تجارب التبريد وجدت القيم الآتية لدرجات الحرارة في عينتين أ ، ب :

أ : ١٠٦,٩ ١٠٦,٣ ١٠٧,٠ ١٠٦,٠ ١٠٤,٩

ب : ١٠٦,٥ ١٠٦,٧ ١٠٦,٨ ١٠٦,١ ١٠٥,٦

هل الفرق بين متوسطى العينتين ذو دلالة عند المستوى ٠,٠٥ ؟
(لتسهيل حساب التباين اطرح ١٠٠ من جميع الأعداد) .

(٥) أخذت عينتان متكافئتان من الأبقار ووضعتا تحت نفس الظروف فيما عدا أنهما أعطيتا نوعين مختلفين من الغذاء ، فكانت الزيادة في الأوزان بالأرطال بعد شهرين كما يلي :

العينة الأولى : ٣٣ ٦٦ ٢٦ ٤٣ ٤٦ ٥٥ ٥٤

العينة الثانية : ٥٣ ٥٣ ٣٧ ٧٣ ٥٨ ٦١ ٣٨

(أ) هل يمكن القول بأن نوعى الغذاء لا يختلفان في متوسط الزيادة في وزن الأبقار ؟

(ب) أوجد فترة ثقة بدرجة ٩٥٪ للفرق بين متوسطى المجتمعين .

(٦) طرح فرض بأن متوسط عدد الزهور الشعاعية ray florets في زهرة ما هو μ = ٥٠ . أخذت عينة حجمها ٥١ من هذه الزهرة فوجد أن متوسط عدد الزهور الشعاعية ٤٤,٣٣ بانحراف معيارى ٥,١٦٧

(أ) اختبر الفرض أن μ = ٥٠

(ب) اختبر الفرض أن μ < ٥٠

(جـ) أوجد فترة ثقة بدرجة ٩٥٪ للدليل μ

(د) أوجد فترة ثقة بدرجة ٩٩٪ للدليل μ

(٦ – ٦ – ٤) اختبار استقلال الأحداث النادرة :

نعلم أنه في توزيع بواسون يتساوى التباين والوسط الحسابي أي أن النسبة بينهما تساوى الواحد الصحيح – راجع الخاصة (٩) من البند (٣ – ٤) . وفي تناولنا للأحداث النادرة في البندين (٣ – ٤ – ٢) ، (٣ – ٤ – ٣) علمنا أن الأحداث النادرة تتوزع بواسونيا بشرط أن تقع مستقلة عن بعضها بمعني أن يكون وقوع هذه الأحداث غير متأثر إلا بالعوامل العشوائية وحدها . وفي كثير من التطبيقات يكون من المرغوب فيه الكشف عن هذه الاستقلالية ثم بحث ما قد يوجد من أنماط نتيجة لعدم توفر هذه الخاصة – راجع البند (٣ – ٤ – ٣) . ولتحقيق هذا الغرض نحصل على عينة عشوائية من تلك الأحداث وليكن حجمها ن ووسطها الحسابي $\bar{س}$ وتباينها $ع^2$ ونضع الفرض الصفرى أن هذه الأحداث تقع عشوائياً مستقلة عن بعضها وبالتالى تتوزع بواسونياً وتكون النسبة بين التباين والوسط الحسابي في المجتمع هى ح = ١ فإذا كان هذا الفرض صحيحاً فإن النسبة $ح_ن = ع^2 / \bar{س}$ المشاهدة في العينة ينبغى أن تكون قريبة من الواحد الصحيح ، أما إذا كان هذا الفرض خاطئاً فإن الفرق بين هذه النسبة والواحد يكون فرقاً ذا دلالة .

ولاختبار صحة الفرض ح = ١ نستخدم الإحصاءة

$$ت = \frac{\frac{ع^2_ن}{\bar{س}} - ١}{\sqrt{\frac{٢}{ن - ١}}} \qquad (١٢)$$

التي يمكن إثبات أن توزيعها يطابق توزيع ت بدرجات حرية عددها ن – ١ . ويجمل بنا أن نستخدم اختباراً ذا جانب واحد لأن المطلوب هنا معرفة ما إذا كانت

ح أكبر من الواحد (أو أقل من الواحد) للمساهمة في تفسير ما قد يوجد من أنماط .

مثال (٦ – ٥) :

اعتبر عينة عشوائية حجمها ٤٠٠ من الأحداث النادرة . اختبر عشوائية (استقلال) هذه الأحداث عند مستوى الدلالة ٠,٠٥ في كل من الحالتين الآتيتين :

(أ) إذا كان الوسط الحسابي للعينة ١,٨ والتباين ١,٩٦٥ .

(ب) إذا كان الوسط الحسابي للعينة ٣,٣ والتباين ٢,١٥ .

الحل :

$$(\text{أ})\ \text{ح}_{\text{ى}} = \text{ع}^{\text{٢}}/\bar{\text{س}} = \frac{\text{١,٩٦٥}}{\text{١,٨}} = \text{١,٠٩٢}$$

الفرض الصفرى $\text{ف}_{\text{٠}}$: ح = ١

الفرض الآخر $\text{ف}_{\text{١}}$: ح > ١

من (١٢) نجد أن :

$$\text{ت}_{\text{ى}} = \frac{\text{١,٠٩٢} - \text{١}}{\sqrt{\frac{\text{٢}}{\text{٣٩٩}}}} = \frac{\text{٠,٠٩٢}}{\text{٠,٠٧١}} = \text{١,٢٩٦}$$

من الجدول (٧) نجد أن $\text{ت}_{\text{٠,١٠}[\text{٣٩٩}]} = \text{١,٦٤٥}$

بما أن ١,٢٩٦ < ١,٦٤٥ لا يكون لدينا دليل ضد الفرض الصفرى ح = ١ ولا يسعنا إلا قبوله . وهذا يعني أن الأحداث تتوزع بواسونياً وبالتالى تقع عشوائياً مستقلة عن بعضها البعض .

$$(\text{ب})\ \text{ح}_{\text{ى}} = \text{ع}^{\text{٢}}/\bar{\text{س}} = \frac{\text{٢,١٥}}{\text{٣,٣}} = \text{٠,٦٥٢}$$

$ف_\cdot : ح = ١$

$ف_١ : ح < ١$

$$ت_ى = \frac{١ - ٠,٦٥٢}{٠,٠٧١} = -٤,٩٠١$$

بما أن $-٤,٩٠١ < -١,٦٤٥$ نرفض الفرض الصفرى عند المستوى ٠,٠٥ ونستنتج أن الأحداث لا تتوزع بواسونياً وبالتالى لا تقع مستقلة عن بعضها البعض ، وأغلب الظن يكون هناك نمط من نوع التنافر .

(٦ - ٧) اختبار χ^2 THE CHI—SQUARE TEST

في كثير من الأبحاث نتناول مجتمعاً صنفت عناصره بحسب خاصة أو صفة ما إلى م من الأقسام المنفصلة $أ_١$ ، $أ_٢$ ، ... ، أم ويكون المطلوب اختبار ما إذا كان هذا المجتمع له نموذج معين من حيث هذه الخاصة ، أى يكون المطلوب اختبار الفرض أن التكرارات النسبية في هذه الأقسام (في المجتمع) لها مقادير معينة $ح_١$ ، $ح_٢$ ، ... ، $ح_م$ على الترتيب .

إذا كان هذا الفرض صحيحاً فإننا لو سحبنا عشوائياً من المجتمع عنصراً واحداً لكان احتمال انتماء هذا العنصر إلى القسم $أ_١$ هو $ح_١$. واحتمال انتمائه إلى القسم $أ_٢$ هو $ح_٢$ ... واحتمال انتمائه إلى القسم $أ_م$ هو $ح_م$. وإذا سحبنا عينة عشوائية حجمها ن فإننا نتوقع أن يكون ن $ح_١$ من عناصرها منتمياً إلى القسم $أ_١$ ، ن $ح_٢$ منها منتمياً إلى القسم $أ_٢$ ... ، ن $ح_م$ منتمياً إلى القسم $أ_م$. نفرض أننا قمنا فعلا بسحب عينة عشوائية حجمها ن ووجدنا أن عدد العناصر التي انتمت إلى الأقسام $أ_١$ ، $أ_٢$ ، ... ، $أ_م$ هي على الترتيب $ك_١$ ، $ك_٢$ ، ... ، $ك_م$. إن هذه الأعداد تسمى بالتكرارات المشاهدة ، أما الأعداد ن $ح_١$ ، ن $ح_٢$ ، ... ، ن $ح_م$ فهي التكرارات النظرية المتوقعة لها وعلى ذلك فإن

اختبار الفرض المذكور يمكن أن يؤسس على المقارنة بين هذين النوعين من التكرارات .

والإحصاءة المناسبة لذلك هى :

$$\text{ص}_{\text{ع}} = \sum_{\text{ر}=1}^{\text{م}} \frac{(\text{ك}_{\text{ر}} - \text{ق}_{\text{ر}})^{2}}{\text{ق}_{\text{ر}}} \qquad \text{حيث } \text{ق}_{\text{ر}} = \text{ح}_{\text{ر}}\,\text{ن} \qquad (13)$$

$$\text{وحيث } \sum \text{ك}_{\text{ر}} = \sum \text{ق}_{\text{ر}} = \text{ن} \qquad (14)$$

التي يمكن إثبات أن توزيعها يقترب من توزيع χ^{2} بدرجات حرية (م − ١) كلما اقتربت ن من اللانهاية .

ولصحة استخدام هذه الإحصاءة ينبغى أن تتحقق الشروط الآتية :

(أ) أن تكون العينة عشوائية .

(ب) ألا يقل حجم العينة عن ٤٠ (لأن توزيع الإحصاءة $\text{ص}_{\text{ع}}$ تقريبى) .

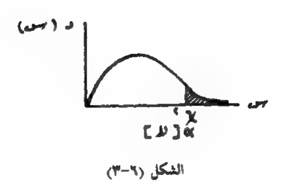
(جـ) ألا تقل أى $\text{ق}_{\text{ر}}$ عن خمسة ، على أنه إذا وجدت قيمة $\text{ق}_{\text{ر}}$ تقل عن خمسة تضم هذه القيمة إلى قيمة مجاورة لها ، ونفعل ذلك لقيم $\text{ك}_{\text{ر}}$ المناظرة .

وجدير بالملاحظة أن (١٣) لا تساوى الصفر إلا إذا كانت $\text{ك}_{\text{ر}} = \text{ق}_{\text{ر}}$ لجميع ر وتزيد قيمتها كلما كبرت الفروق بين $\text{ك}_{\text{ر}}$ ، $\text{ق}_{\text{ر}}$.

إن الفرض الصفرى $\text{ف}_{\cdot}$ هنا هو عدم وجود فرق بين التكرارات المشاهدة $\text{ك}_{\text{ر}}$ والتكرارات $\text{ق}_{\text{ر}}$ المتوقعة لها . أما الفرض الآخر ف_{1} فهو أن هنالك تفاوتاً بين هذين النوعين من التكرارات يجعل القيمة المشاهدة للإحصاءة (١٣) كبيرة كبراً ذا دلالة . وعلى ذلك تأخذ المنطقة الحرجة الصورة الآتية :

$$\text{م} = \left\{ \chi^{2} : \chi^{2} > \chi^{2}_{\alpha[\nu]} \right\}$$

حيث ν ترمز إلى درجات الحرية .



الشكل (٦-٣)

حساب درجات الحرية :

نعلم أن توزيع χ^2 يتوقف كلية على عدد درجات الحرية ν . ولايجاد هذا العدد نحسب عدد المقارنات المستقلة بين $ك_ر$ ، $ق_ر$ (حيث ر = ١ ، ٢ ، . . . ، م) ونظراً لأن حجم العينة ن هو مقدار ثابت معروف مقدماً فإن التكرارات $ك_ر$ (أو $ق_ر$) لا تكون جميعها مستقلة إذ نستطيع إذا عرفنا (م − ١) منها أن نعرف من المتساوية (١٤) التكرار الباقي . وعلى هذا فإن عدد المقارنات المستقلة هو م − ١ . فإذا كانت القيمة المشاهدة للإحصاءة (١٣) قد حسبت اعتماداً على حجم العينة فقط فإن ν = م − ١ .

وكقاعدة عامة كل قيد خطي تتقيد به القيم الداخلة في تركيب الإحصاءة (١٣) ينقص واحداً من درجات الحرية ، ولقد كان ثبات حجم العينة هو أحد هذه القيود . وإذا كنا قد احتجنا لحساب القيم المتوقعة إلى واحد أو أكثر من أدلة المجتمع المجهولة واضطررنا إلى تقديرها من العينة فإننا نكون قد قيدنا أنفسنا بالعينة التي بأيدينا وبالتالى فإن كلا من هذه التقديريات يشكل قيداً على التوزيع ينقص واحداً من درجات الحرية . وكطريقة إجرائية سهلة ، نحسب عدد درجات الحرية كالآتي :

ν = عدد المقارنات المستقلة − عدد الأدلة التي قدرت من العينة (١٥)

فيما يلي ثلاثة تطبيقات تستخدم اختبار χ^2 بالصيغة المبينة في (١٣) .

(٦ – ٧ – ١) اختبار فرض عن توزيع مجتمع :

مثال (٦ – ٦) :

في أحد تجارب مندل الشهيرة في تهجين النباتات نتج من نبات البسلة ما يلي : ٣١٥ نباتاً مستديراً أصفرا ، ١٠٨ مستديراً أخضرا ، ١٠١ مجعدا أصفرا ، ٣٢ مجعداً أخضرا . وطبقاً لنظرية مندل يجب أن تكون أعداد هذه الأنواع متناسبة مع ٩ : ٣ : ٣ : ١ . هل هناك دليل يدعو إلى الشك في نظرية مندل من واقع هذه البيانات ؟ استخدم مستوى الدلالة ٠,٠٥ .

الحـــل :

لدينا ٣١٥ + ١٠٨ + ١٠١ + ٣٢ = ٥٥٦ نباتاً (حجم العينة) .

إذا كانت نظرية مندل صحيحة ، تتوزع هذه النباتات بالنسب ٩ : ٣ : ٣ : ١ كالآتي :

مستدير أصفر = ٥٥٦ × $\frac{٩}{١٦}$ = ٣١٢,٧٥

مستدير أخضر = ٥٥٦ × $\frac{٣}{١٦}$ = ١٠٤,٢٥

مجعد أصفر = ٥٥٦ × $\frac{٣}{١٦}$ = ١٠٤,٢٥

مجعد أخضر = ٥٥٦ × $\frac{١}{١٦}$ = ٣٤,٧٥

التكرارات المشاهدة $ل_ر$:	٣١٥	١٠٨	١٠١	٣٢
التكرارات المتوقعة $ه_ر$:	٣١٢,٧٥	١٠٤,٢٥	١٠٤,٢٥	٣٤,٧٥

الفرض الصفرى $ف_٠$: نظرية مندل صحيحة .

الفرض الآخر $ف_١$: نظرية مندل غير صحيحة وبالتالى تكون القيمة المشاهدة للإحصاء (١٣) كبيرة كبراً ذا دلالة .

$$\chi^{٢}_{م} = \frac{(٣١٥-٣١٢,٧٥)^{٢}}{٣١٢,٧٥} + \frac{(١٠٨-١٠٤,٢٥)^{٢}}{١٠٤,٢٥} + \frac{(١٠١-١٠٤,٢٥)^{٢}}{١٠٤,٢٥} + \frac{(٣٢-٣٤,٧٥)^{٢}}{٣٤,٧٥} = ٠,٤٧$$

حيث ν = ٤ − ١ =٣

من جدول χ^2 نجد أن $\chi^2_{٠,٠٥[٣]}$ = ٧,٨٢

العدد ٠,٤٧ يقل كثيراً عن ٧,٨٢ (لا يقع في منطقة الرفض) وإذن لا يمكن رفض الفرض الصفرى عند مستوى الدلالة ٠,٠٥ وبالتالى نقبله ونقرر أنه لا يوجد دليل يدعو إلى الشك في نظرية مندل وأن نتيجة هذه التجربة تدعم هذه النظرية .

مثال (٦ − ٧) :

المفروض في جداول الأعداد العشوائية أن تظهر الأرقام ٠ ، ١ ، ٢ ، ٠٠٠ ، ٩ باحتمالات متساوية . بفحص إحدى صفحات أحد هذه الجداول وجد التوزيع الآتي :

الرقم	٠	١	٢	٣	٤	٥	٦	٧	٨	٩
التكرار	١٧	٣١	٢٩	١٨	١٤	٢٠	٣٥	٣٠	٢٠	٣٦

اختبر الفرض أن الأرقام العشرة تظهر باحتمالات متساوية مستخدماً مستوى الدلالة ٠,٠١

الحـــل :

مجموع الأرقام التي بالصفحة = ١٧ + ٣١ + ٠٠٠ + ٣٦ = ٢٥٠ رقماً

إذا كانت الأرقام العشرة تظهر باحتمالات متساوية فإن كلا منها يجب أن يظهر ٢٥٠ ÷ ١٠ = ٢٥ مرة .

أي أن التكرارات المتوقعة على أساس صحة هذا الفرض هي ٢٥ لكل من الأعداد العشرة .

$$\chi^2_{ع} = \frac{(١٧-٢٥)^2}{٢٥} + \frac{(٣١-٢٥)^2}{٢٥} + \ldots\ldots + \frac{(٣٦-٢٥)^2}{٢٥}$$

$= ٢٣,٣$ حيث $\nu = ١٠ - ١ = ٩$

القيمة الحرجة $\chi^2_{٠,٠١[٩]} = ٢١,٦٦٦$

بما أن ٢٣,٣ أكبر من ٢١,٦٦٦ نرفض الفرض الصفرى عن تساوى احتمالات ظهور الأرقام العشرة عند مستوى الدلالة ٠,٠١ ويتضمن هذا أن هناك شك في دقة هذا الجدول .

(٦ - ٧ - ٢) اختبار حسن المطابقة

TEST OF GOODNESS OF FIT

يقصد بهذا الاختبار تحديد مدى ملاءمة توزيع نظرى معروف مثل ذى الحدين وبواسون والمعتدل ، لتوزيع تكرارى مشاهد في عينة ، بهدف التحقق من صلاحيته كتوزيع للمجتمع الذى أخذت منه العينة .

مثال (٦ - ٨) :

اختبر الفرض القائل أن إنجاب الذكور وإنجاب الإناث متساويا الاحتمال من واقع البيانات المعطاة في المثال (٣ - ٥) عن العائلات ذوات الأربعة الأطفال وهى :

عدد الذكور في العائلة	$س_ر$:	٠	١	٢	٣	٤
عـــدد العـــائلات	$ك_ر$:	١٣	٤٣	٩٢	١١٢	٦٠

الحل :

على فرض صحة الفرض الصغرى أن إنجاب الذكور وإنجاب الإناث متساويا الاحتمال يكون احتمال إنجاب الذكور ح = $\frac{1}{٢}$ ويكون توزيع عدد الذكور في العائلات

ذوات الأربعة الأطفال هو توزيع ذى الحدين دليلاه ٤ ، $\frac{1}{2}$. وقد وجدنا في المثال (٣ — ٥) أن هذا التوزيع هو :

عدد الذكور في العائلة $س_ر$:	٠	١	٢	٣	٤
عـــدد العـــائلات $ق_ر$:	٢٠	٨٠	١٢٠	٨٠	٢٠

$$\chi^2_{ر} = \frac{(١٣-٢٠)^2}{٢٠} + \frac{(٤٣-٨٠)^2}{٨٠} + \frac{(٩٢-١٢٠)^2}{١٢٠} + \frac{(١١٢-٨٠)^2}{٨٠} + \frac{(٦٠-٢٠)^2}{٢٠}$$

$= ١١٨,٨٩$ حيث $\nu = ٥ - ١ = ٤$

$$\chi^2_{٠,٠١[٤]} = ١٣,٢٧٧$$

بما أن ١١٨,٨٩ > ١٣,٢٧٧ نرفض صحة الفرض أن احتمال إنجاب الذكور يساوى احتمال إنجاب الإناث ، وذلك عند مستوى الدلالة ٠,٠١

مثال (٦ — ٩) اختبار الاعتدالية : TEST OF NORMALITY

اختبر ما إذا كان من الممكن اعتبار العينة المعطاة في المثال (٤ — ٤) مأخوذة من مجتمع معتدل .

الحل :

في ذلك المثال وفقنا توزيعاً معتدلا للتوزيع المعطى ووجدنا ما يلى :

التكرارات المشاهدة $ك_ر$	٦	٦	١١	١٤	١٦	١٥	٨
التكرارات المتوقعة $ق_ر$	٦,٨١	٦,٥٤	٩,٦١	١٢,٩٨	١٣,٦٦	١٥,٥٧	١٢,٤٧

التكرارات المشاهدة $ك_ر$:	١٠	٨	٦
التكرارات المتوقعة $ق_ر$:	٩,٤٤	٦,٣٧	٦,٥٥

$$\chi^2_{م} = \frac{(٦ - ٦,٨١)^2}{٦,٨١} + \frac{(٦ - ٦,٥٤)^2}{٦,٥٤} + \ldots\ldots + \frac{(٨ - ٦,٣٧)^2}{٦,٣٧} + \frac{(٦ - ٦,٥٥)^2}{٦,٥٥} = ٢,٩٤٢٧$$

بما أننا كنا قد قدرنا من العينة اثنين من أدلة المجتمع هما الوسط الحسابي والانحراف المعياری فإن عدد درجات الحرية ν = (١٠ − ١) − ٢ = ٧ .

$$\chi^2_{٠,٠٥\ [٧]} = ١٤,٠٦٧$$

بما أن ٢,٩٤٢٧ > ١٤,٠٦٧ لا نستطيع رفض الفرض الصفرى عن اعتدالية المجتمع عند مستوى الدلالة ٠,٠٥ .

ملاحظة (١) :

إن عيب اختبار χ^2 لحسن المطابقة أنه لا يهتم بإشارات الفروق بين التكرارات المشاهدة ك‌م والتكرارات النظرية ق‌م (لأنه يأخذ مربعات هذه الفروق) ففي توفيق توزيع معتدل مثلا قد يحدث أن تكون قيم ق‌م أصغر من قيم ك‌م في الجزء الأوسط من التوزيع وتكون أكبر منها في الطرفين وهذا يعني أن التوزيع المشاهد أكثر انبساطاً من التوزيع المعتدل أى يأخذ شكلا يختلف عن شكل التوزيع المعتدل . ومع هذا قد تكون قيمة $\chi^2_{م}$ ليست بذات دلالة وتدعونا إلى اعتبار أن المجتمع معتدل . ولهذا لا يكون الاختبار مناسباً إلا إذا لم يكن هناك نمط معين للفروق بين ق‌م ، ك‌م .

(٦ − ٧ − ٣) اختبار استقلال خاصتين :

TEST OF INDEPENDENCE

في بعض الدراسات نتناول مجتمعاً صنفت عناصره بحسب خاصة ما إلى م من الأقسام المنفصلة $ا_١$ ، $ا_٢$ ، ... ، $ا_م$ وصنفت من ناحية ثانية بحسب خاصة أخرى إلى هـ من الأقسام المنفصلة $ب_١$ ، $ب_٢$ ، ... ، $ب_هـ$ ويكون المراد اختبار ما إذا كانت هاتان الخاصتان مستقلتين ، بمعني أن يكون توزيع إحداهما غير متأثر بالأخرى .

في هذه الحال تؤخذ عينة عشوائية وتدرس من حيث عدد العناصر التي تظهر في أقسام الخاصتين وتوضع التكرارات الناتجة بحيث تقترن التكرارات في أقسام الخاصة الأولى بالتكرارات في أقسام الخاصة الثانية فيما يسمى بجدول الاقتران contingency table وهو عبارة عن مصفوفة ذات م صفا تمثل أقسام الخاصة الأولى ، هـ عموداً تمثل أقسام الخاصة الثانية ويوصف الجدول حينئذ بأنه جدول اقتران م × هـ . نضع الفرض الصفرى أن الخاصتين مستقلتان ، وعلى أساس صحة هذا الفرض نحسب التكرارات النظرية المناظرة للتكرارات المشاهدة ونضعها في جدول اقتران م × هـ أيضا . بمقارنة التكرارات المشاهدة بالتكرارات النظرية نستطيع الحكم على استقلال الخاصتين بواسطة الإحصاءة المعرفة في (١٣) .

مثال (٦ – ١٠) :

في تجربة عن الرجال في الأعمار بين ٦٠ ، ٦٤ عاماً صنف الرجال من حيث خاصتي التدخين والوفاة ، وقسمت خاصة التدخين إلى قسمين هما « يدخن ولا يدخن » وقسمت خاصة الوفاة إلى قسمين : رجال لا يزالون على قيد الحياة ورجال توفوا في بحر ٦ سنوات من بدء التجربة . في عينة من ١٤٦٩ رجلا جمعت البيانات في جدول الاقتران ٢ × ٢ الآتي :

الوفاة	عادة التدخين		المجموع
	يدخن	لا يدخن	
توفي	٥٤	١١٧	١٧١
حي	٣٤٨	٩٥٠	١٢٩٨
المجموع	٤٠٢	١٠٦٧	١٤٦٩

هل هذه البيانات تشير إلى أن الوفاة مستقلة عن عادة التدخين ؟

الحل :

ف : الوفاة وعادة التدخين خاصتان مستقلتان ، أو لا توجد علاقة بين الخاصتين .

إذا كان هذا الفرض صحيحاً تكون نسبة المتوفين إلى الأحياء واحدة في المجتمع سواء للمدخنين أو لغير المدخنين . ونظراً لأن هذه النسبة غير معروفة نضطر إلى تقديرها من العينة ، ومن المعقول أن نأخذ النسبة التي ظهرت في العينة بين العدد الكلي للمتوفين والعدد الكلي للأحياء وهي ١٧١ : ١٢٩٨ . وباستخدام هذه النسبة نحصل على التكرارات النظرية في الخلايا الأربع وذلك بتقسيم كل من عدد الذين يدخنون وهو ٤٠٢ وعدد الذين لا يدخنون وهو ١٠٦٧ بهذه النسبة فمثلا :

التكرار النظرى لعدد المتوفين من المدخنين = $٤٠٢ \times \frac{١٧١}{١٤٦٩} = ٤٦,٨$ رجلا

التكرار النظرى لعدد الأحياء من المدخنين = $٤٠٢ \times \frac{١٢٩٨}{١٤٦٩} = ٣٥٥,٢$ رجلا

وبالمثل بالنسبة لغير المدخنين . وبذلك نحصل على جدول التكرارات النظرية الآتي :

الوفاة	عادة التدخين		المجموع
	يدخن	لا يدخن	
توفي	٤٦,٨	١٢٤,٢	١٧١
حي	٣٥٥,٢	٩٤٢,٨	١٢٩٨
المجموع	٤٠٢	١٠٦٧	١٤٦٩

$$\chi^{٢} = \frac{(٥٤-٤٦,٨)^{٢}}{٤٦,٨} + \ldots + \ldots + \frac{(٩٥٠-٩٤٢,٨)^{٢}}{٩٤٢,٨} = ١,٧٣$$

بحسب عدد درجات الحرية كالآتي . لدينا ٤ مقارنات ، وبما أن عدد المدخنين ثابت (وهو ٤٠٢) وكذلك عدد غير المدخنين (وهو ١٠٦٧) فإن درجات الحرية تنقص ٢ ، وبما أننا قدرنا دليل واحد للمجتمع من العينة وهو النسبة ١٧١ : ١٢٩٨ فإن درجات الحرية تنقص واحداً آخر .

إذن ν = (٤−٢) −١ = ١

$\chi^{٢}_{٠,٠٥[١]}$ = ٣,٨٤١

بما أن ١,٧٣ < ٣,٨٤١ لا نستطيع رفض الفرض الصفرى عند المستوى ٠,٠٥ ويمكننا القول بأنه في حدود مجموعة العمر التي درست ، عادة التدخين والممات مستقلتان ، أى أن التدخين لا يؤثر في الوفاة في حدود هذه التجربة .

مثال (٦ − ١١) :

جدول الاقتران ٢ × ٣ الآتي يحمل التكرارات المشاهدة في عينة عشوائية من ٩٩ طفلا حديث الولادة من حيث طول الطفل وطول محيط رأسه بالسنتيمترات ساعة الولادة . ابحث استقلال طول المولود وطول محيط رأسه .

محيط الرأس	طول المولود			المجموع
	٤٧−٤٩	٥٠−٥٢	٥٣−٥٥	
٣٢−٣٥	٤٠	٣٦	٢	٧٨
٣٦−٣٩	٠	١٤	٧	٢١
المجموع	٤٠	٥٠	٩	٩٩

الحل :

كما في المثال السابق ، لو كان طول المولود ومحيط رأسه مستقلين لتوزعت أعداد الأطفال في كل من الأقسام الثلاثة للأطوال بنفس النسبة . ونقدر هذه النسبة من العينة على أنها ٧٨ : ٢١ أى ينبغى أن نقسم كلا من الأعداد ٤٠ ، ٥٠ ، ٩ بالنسبة ٧٨ : ٢١ لنحصل على جدول التكرارات النظرية الآتي :

محيط الرأس	طول المولود			المجموع
	٤٧–٤٩	٥٠–٥٢	٥٣–٥٥	
٣٢–٣٥	٣١,٥	٣٩,٤	٧,١	٧٨
٣٦–٣٩	٨,٥	١٠,٦	١,٩	٢١
المجموع	٤٠	٥٠	٩	٩٩

$$\chi^2_{م} = \frac{(٤٠-٣١,٥)^2}{٣١,٥} + \ldots + \frac{(٧-١,٩)^2}{١,٩} = ٢٩,٣٧٧$$

$$\nu = (٣-٦) - ١ = ٢$$

$$\chi^2_{٠,٠٥[٢]} = ٥,٩٩١$$

نرفض الفرض الصفرى عن استقلال طول المولود وطول محيط رأسه عند مستوى الدلالة ٠,٠٥ ونحكم بوجود علاقة بين هاتين الخاصتين .

ملاحظة (١)

(أ) إذا كانت هناك علاقة موجبة بين الخاصتين تظهر تكرارات كبيرة نسبيا فى خلايا القطر الرئيسى أى فى الخلايا العلوية اليمنى وفى الوسط وفى الخلايا السفلية اليسرى كما فى هذا المثال .

(ب) وإذا كانت هناك علاقة سالبه تظهر تكرارات كبيرة نسبيا فى الخلايا السفلية اليمنى وفى الوسط وفى الخلايا العلوية اليسرى .

(جـ) أما إذا كان هناك علاقة صغيرة أو لا توجد علاقة فإن التكرارات تميل إلى أن تكون متناسبة الكثافة بمعنى أن يكون توزيع التكرارات بنفس النسبة التى تظهر فى المجاميع الهامشية .

ملاحظة (٢) :

هناك قاعدة سهلة لإيجاد عدد درجات الحرية في أى جدول اقتران م × هـ حيث م > ١ ، هـ > ١ وهى :

$$\nu = (م - ١)(هـ - ١) \qquad (١٦)$$

وتفسير هذه القاعدة أنه مادامت المجاميع الهامشية للأعمدة ثابتة فإننا في ملأ خانات جدول الاقتران يمكننا استنتاج واحد من الأعداد التي بأى عمود من الـ هـ − ١ الأعداد الأخرى التي بهذا العمود ، وبالمثل يمكن استنتاج واحد من الأعداد التي بأى صف من الـ م − ١ الأعداد الأخرى التي بهذا الصف . وبهذا نكون أحراراً فقط في ملأ (م − ١) (هـ − ١) من خلايا الجدول .

ملاحظة (٣) :

هناك أيضاً قاعدة سهلة لحساب التكرار النظرى المناظر لتكرار مشاهد في أى خلية ، وذلك بضرب المجموعين الهامشيين للصف والعمود الذى تقع فيه الخلية

ثم قسمة الناتج على المجموع الكلى للتكرارات (حجم العينة) ، فمثلا في المثال (٦ – ١١) الأخير :

$$\text{التكرار النظرى المناظر للعدد ٣٦ هو } \frac{٧٨ \times ٥٠}{٩٩} = ٣٩,٤$$

$$\text{، التكرار النظرى المناظر للعدد ٧ هو } \frac{٢١ \times ٩}{٩٩} = ١,٩$$

(٦ – ٧ – ٤) فترات الثقة لتباين مجتمع معتدل :

في البند (٦ – ٦ – ٢) أوجدنا صيغة لفترات الثقة للوسط الحسابي μ لمجتمع معتدل واستخدمنا لذلك توزيع ت . أما فترات الثقة لتباين مجتمع معتدل فتحتاج لاستخدام توزيع χ^2 ، إذ يمكن إثبات أنه على أساس عينة عشوائية حجمها ن وتباينها $ع^2$ مأخوذة من مجتمع معتدل تباينه σ^2 فإن المتباينة :

$$\frac{(ن - ١)ع^2}{\chi^2_{(١-\frac{\alpha}{٢})}[ن-١]} < \sigma^2 < \frac{(ن - ١)ع^2}{\chi^2_{\frac{\alpha}{٢}}[ن-١]} \quad (١٧)$$

تشكل فترة ثقة بدرجة (١ – α) لتباين المجتمع . وليس من الضرورى هنا أن تكون العينة كبيرة الحجم كما هو الحال في التطبيقات سابقة الذكر لأن التوزيع الذى بنيت عليه هذه الفترة هو توزيع مضبوط طالما كان المجتمع معتدلا .

مثال (٦ – ١٢) :

أخذت عينة عشوائية حجمها ١٢ من مجتمع معتدل فوجد أن تباينها ٩,٧٣ . أوجد فترة ثقة بدرجة ٩٥٪ لتباين المجتمع .

الحل :

لدينا $\alpha = ٠,٠٥$ ‏ $\frac{\alpha}{٢} = ٠,٠٢٥$ ‏ $١ - \frac{\alpha}{٢} = ٠,٩٧٥$ ‏ $ن = ١٢$

من جدول $\chi^٢$ نجد أن $\chi^٢_{\frac{\alpha}{٢}\,[ن-١]} = \chi^٢_{٠,٠٢٥\,[١١]} = ٢١,٩$

، $\chi^٢_{(١-\frac{\alpha}{٢})\,[ن-١]} = \chi^٢_{٠,٩٧٥\,[١١]} = ٣,٨٢$

بالتعويض فى (١٧) نجد أن :

الحد الأدنى للفترة = $\frac{١١ \times ٩,٧٣}{٢١,٩} = ٤,٨٨$

الحد الأعلى للفترة = $\frac{١١ \times ٩,٧٣}{٣,٨٢} = ٢٨,٠١٨$

إذن الفترة (٤,٨٨ ، ٢٨,٠١٨) هى فترة ثقة بدرجة ٩٥٪ لتباين المجتمع .

تمارين (٦ – ٢)

(فى كل المسائل الآتية اذكر الشروط اللازم توفرها لصحة الحل .)

(١) حسب نظرية مندل عن تهجين النباتات تكون نسبة نبات البسلة الأخضر إلى نبات البسلة الأصفر ٣ : ١ . اختبر هذه النظرية على ضوء البيانات المشاهدة الآتية :

(أ) وجد في عينة عشوائية أن هناك ٣٥٥ نباتاً أخضر ، ١٢٣ نباتاً أصفر .

(ب) وجد في عينة عشوائية أن هناك ٤٢٨ نباتاً أخضر ، ١٥٢ نباتاً أصفر .

(٢) وضعت خمس مصايد فيران في أماكن مختلفة من غابة ما . في فترة ثلاثة أشهر سجلت أعداد الفيران المصيدة كالآتى :

المصيدة :	أ	ب	حـ	د	هـ
عدد الفيران :	٢٣	٧	٢٥	١٩	٢١

اختبر الفرض أنه لا يوجد فرق بين المصايد الخمس في عدد ما تصطاده من الفيران .

(٣) اختبر الفرض أن عدد المواليد اليومية في مجتمع ما ثابت خلال شهور السنة مستخدماً البيانات الآتية :

الشهر :	يناير	فبراير	مارس	ابريل	مايو	يونيو	يوليو	أغسطس	سبتمبر	أكتوبر	نوفمبر	ديسمبر
عدد المواليد :	٨٠	٧٨	٨٦	٨٢	٨٣	٧٨	٧٩	٧٦	٧٨	٧٦	٧٢	٧٦

(٤) لمعرفة ما إذا كان مجتمع ما يفضل نوعاً أ من سلعة ما (معجون أسنان مثلا) على نوع آخر ب أجرى استفتاء على عينة من ٣٠٠ شخص فوجد أن ١٦٨ شخصاً منهم يفضلون النوع أ وأن ١٣٢ شخصاً يفضلون النوع ب . فهل هذه النتيجة تعني أن المجتمع يفضل النوع أ ؟

(٥) في ٣٦٠ رمية لزهرتين من النرد ظهر ما مجموعة سبعة ٧٤ مرة وظهر ما مجموعة إحدى عشر ٢٤ مرة . اختبر الفرض أن الزهرتين غير متحيزتين .

(٦) في المثال (٣ — ٩) عن توزيع خلايا الخميرة وفقنا توزيع بواسون لتوزيع تكرارى مشاهد في عينة . استخدم المستوى ٠,٠٥ لاختبار حسن المطابقة .

(٧) أخذ جوالان من حبوب الشعير لاختبار تأثير معالجة حرارية معينة على حيوية الحبوب وقد ترك الجوال أ دون معالجة (مجموعة مراقبة) وأعطيت المعالجة للجوال ب ثم أخذت عينة من ٨٠ حبة من كل جوال وفحصت بطريقة ما من حيث حيوية الحبوب فكانت النتيجة كما يلي :

	يحتوى على مقومات الحياة	لا يحتوى على مقومات الحياة
الجوال أ :	٦٤	١٦
الجوال ب :	٣٤	٤٦

هل هناك فرق ذو دلالة بين الجوالين نتيجة للمعالجة الحرارية ؟ أى هل حيوية الحبوب مستقلة عن المعالجة الحرارية ؟

(٨) أخذت مجموعتان متكافئتان أ ، ب كل منهما تتكون من ١٠٠ شخص مريض بمرض معين وأعطى دواء للمجموعة الأولى ، ولم يعط للمجموعة الثانية (مجموعة المراقبة) فوجد أن عدد الذين شفوا من المجموعتين أ ، ب هما ٧٥ ، ٦٥ على الترتيب . أختبر الفرض أن هذا الدواء يساعد على الشفاء من المرض .

(٩) في عينة من خمس فراشات وجد أن تباين طول الجناح ١٣,٥٢ . إذا فرض أن مجتمع أجنحة هذه الفراشات معتدل (بالنسبة لطول الجناح) فأوجد فترة ثقة بدرجة ٩٥٪ لتباين هذا المجتمع .

(١٠) فى دراسة عن نوع من الضفادع وجد في عينة من ٣٩ ضفدعا أن متوسط فترة نداءات الذكور ١٨٩ وحدة بانحراف معيارى ٣٢ وحدة .

(أ) أوجد فترة ثقة بدرجة ٩٥٪ لمتوسط فترة النداء في المجتمع .

(ب) أوجد فترة ثقة بدرجة ٩٥٪ لتباين فترة النداء في المجتمع .

(١١) بجدول الاقتران الآتى التوزيع المشترك لمجموعة من ٩٥٠ طالبا من حيث خاصتى الذكاء والتغذية . هل تشير هذه البيانات إلى أن ذكاء الطلاب مستقل عن التغذية ؟

نوع التغذية	نسبة الذكاء				المجموع
	أقل من ٨٠	٨٠–٨٩	٩٠–٩٩	١٠٠ فأكثر	
تغذية حسنة	٢٤٥	٢٢٨	١٧٧	٢١٩	٨٦٩
سوء تغذية	٣١	٢٧	١٣	١٠	٨١
المجموع	٢٧٦	٢٥٥	١٩٠	٢٢٩	٩٥٠

(٦ – ٧ – ٥) خاصة الجمع لتوزيعات χ^2 المستقلة :

يمكن إثبات النظرية الآتية : « إذا كان لمتغير ما توزيع χ^2 بدرجات حرية ν_1 وكان لمتغير آخر **مستقل** عن المتغير الأول توزيع χ^2 بدرجات حرية ν_2 فإن المتغير الذى ينشأ من ضم هذين المتغيرين معا يكون له توزيع χ^2 بدرجات حرية $\nu_1 + \nu_2$ » وتمتد هذه النظرية لأى عدد منتهى من المتغيرات المستقلة .

مثال (٦ – ١٣) :

فى دراسة عن تأثير التطعيم بمصل ما فى الوقاية من مرض ما أخذت مجموعتان متكافئتان من الأطفال طعمت إحداهما بالمصل ولم تطعم الأخرى . وبعد فترة محددة حسب عدد الذين أصيبوا والذين لم يصابوا بالمرض فى كلتا المجموعتين ووضعت النتائج فى جدول ٢ × ٢ ووجد أن χ^2 = ٤,١ بدرجة حرية واحدة . كررت نفس الدراسة مرتين أخريتين بشكل مستقل (فى مكانين آخرين) ووجد أن قيمتى χ^2 هما ٣,٩ ، ٤,٢ بدرجة حرية واحدة لكل منهما . نلاحظ ما يلى :

(أولا) إذا أخذنا كلا من هذه التجارب الثلاث على حدة نجد أن كلا منها يشير إلى فرق ذى دلالة بين المجموعة التى طعمت بالمصل والمجموعة الضابطة عند المستوى ٠,٠٥ وذلك لأن القيمة الحرجة $\chi^2_{٠,٠٥[١]}$ = ٣,٨٤١ وهذا يعنى أن المصل كان له تأثير فى الحماية من المرض .

(ثانيا) إذا ضممنا نتائج التجارب الثلاث معا ، وهذا جائز لأنها مستقلة ، نجد أن :

المجموع = ٤,١ + ٣,٩ + ٤,٢ = ١٢,٢ بدرجات حرية عددها ثلاث ونظرا لأن القيمة الحرجة $\chi^2_{٠,٠٥[٣]}$ = ٧,٨١٥ فإن هذه النتيجة تدعم كلا من النتائج الثلاث السابقة وتؤكد أن الفروق بين أزواج المجموعات لم تكن بالصدفة .

(٦ – ٨) اختبار ف : THE F - TEST

إن الأهمية الرئيسية لهذا الاختبار هو قدرته على اختبار تساوى تبايني مجتمعين معتدلين σ^{2}_{1} ، σ^{2}_{2} عن طريق تبايني عينتين عشوائيتين مستقلتين ع$^{2}_{1}$ ، ع$^{2}_{2}$ مأخوذتين منهما . وقد رأينا أن اختبار ت عن دلالة الفرق بين متوسطى عينتين بالصورة المبينة بالبند (٦ ـــ ٦ ـــ ٣) يستلزم التأكد من تساوى تبايني المجتمعين ، ولذا ينبغى إجراء اختبار ف بشكل روتيني قبل تطبيق هذا الاختبار . ومن التطبيقات الهامة التي يستخدم فيها اختبار ف ذلك التطبيق المسمى بتحليل التباين الذى سنتناوله في الفصل الثامن حيث نحتاج أيضاً إلى اختبار تساوى تباينات المجتمعات .

ويبني هذا الاختبار على الإحصاءة :

$$\text{ف}_{\cdot} = \frac{\text{ع}^{2}_{1}}{\text{ع}^{2}_{2}} \qquad (18)$$

التي تسمى بنسبة التباين variance ratio والتي يتطابق توزيعها مع توزيع المتغير ف السابق دراسته بدرجتي حرية ν_{1} = ن$_{1}$ −١ ، ν_{2} = ن$_{2}$ −١تحت الافتراضات الآتية :

(أ) أن تكون كل من العينتين عشوائية بسيطة .

(ب) أن تكون العينتان مستقلتين .

(حـ) أن تكون العينتان مأخوذتين من مجتمع معتدل واحد أو من مجتمعين معتدلين لهما نفس التباين .

ملاحظة :

لاستخدام جدول ف للقيم الحرجة يجب أن توضع نسبة التباين بحيث تكون

أكبر من الواحد أى بحيث تكون $ع_1^2 > ع_2^2$ مع ملاحظة أن ν_1 تحدد العمود ، ν_2 تحدد الصف .

مثال (٦ – ١٤) :

أخذت عينتان عشوائيتان حجماهما ٥ ، ٦ من الإناث والذكور من حشرة ما وحسبت مدة البقاء على الحياة لكل حشرة بعد حرمانها من الطعام والشراب فوجد أن تباين الأولى ٢٥,٧ وتباين الثانية ٧,٠٦ على أساس أن المجتمعين معتدلين :

(أولا) اختبر عند مستوى الدلالة ٠,٠١ أن تباين مجتمع الإناث أكبر من تباين مجتمع الذكور .

(ثانيا) اختبر عند مستوى الدلالة ٠,٠٥ أن المجتمعين متساويا التباين .

الحل :

$$ف = \frac{٢٥,٧}{٧,٠٦} = ٣,٦٤ \text{ بدرجتي حرية } ٤ ، ٥$$

(أولا) الفرض الصفرى ف. : $\sigma_1^2 = \sigma_2^2$

الفرض البديل ف$_1$ $\sigma_1^2 > \sigma_2^2$ (اختبار ذو جانب واحد)

من الجدول (٩) : $ف_{٠,٠١\,[٤،٥]} = ١١,٤$

بما أن ٣,٦٤ < ١١,٤ لا نستطيع رفض ف. عند المستوى ٠,٠١

(ثانياً) الفرض الصفرى ف. : $\sigma_1^2 = \sigma_2^2$

الفرض الآخر ف$_1$: $\sigma_1^2 \neq \sigma_2^2$ (اختبار ذو جانبين مساحة كل منهما ٠,٠٢٥)

نريد إيجاد $ف_١$ ، $ف_٢$ حيث ل (ف > $ف_١$) = ٠,٠٢٥

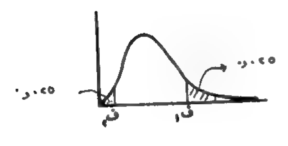
، ل (ف > $ف_٢$) = ٠,٩٧٥ مع ملاحظة أن الاحتمال ٠,٩٧٥ لا يوجد بالجدول .

من الجدول $ف_{٠,٠٢٥ [٥ ، ٤]}$ = ٧,٣٩

إذن القيمة الحرجة اليمني $ف_١$ = ٧,٣٩

المتساوية ل (ف > $ف_٢$) = ٠,٩٧٥ تكافىء

المتساوية ل $\left(\frac{١}{ف} > \frac{١}{ف_٢} \right)$ = ٠,٠٢٥



الشكل (٦-٤)

من الجدول $ف_{٠,٠٢٥ [٤ ، ٥]}$ = ٩,٣٦ راجع ملاحظة البند (٥-٣)

∴ $\frac{١}{ف_٢}$ = ٩,٣٦

∴ القيمة الحرجة اليسرى $ف_٢$ = ١ ÷ ٩,٣٦ = ٠,١٠٧

بما أن ٧٣,٩ > ٣,٦٤ > ٠,١٠٧ لا نستطيع رفض $ف_٠$ عند المستوى ٠,٠٥

أى أننا نقبل أن $\sigma_١^٢ = \sigma_٢^٢$ عند هذا المستوى .

(٦ – ٩) فترات الثقة للنسبة في مجتمع :

نفرض أن نسبة وقوع حدث أ في مجتمع ما هو عدد مجهول ح ، فمثلا قد تعبر ح عن نسبة ظهور زهور حمراء من نوع من البذور ، أو نسبة الإصابة بمرض ما في نوع من الماشية ، أو نسبة النيكوتين في نوع من السجائر . يمكننا تقدير النسبة ح من واقع عينة عشوائية كبيرة وذلك بأخذ النسبة المشاهدة ر بين عدد مرات وقوع الحدث أ وعدد وحدات العينة ، وتزداد ثقتنا في التقدير ر كلما زاد حجم العينة . وإذا أردنا إيجاد صيغة لفترات الثقة للنسبة ح فإننا نستخدم التقدير ر كأساس لبناء هذه الصيغة وبحسب المنطق الآتي :

نعلم من البند (٣ – ٣) أنه تحت شروط عشوائية العينة وثبات النسبة ح واستقلال الأحداث يكون للمتغير العشوائي س الذى يعبر عن عدد مرات وقوع الحدث أ في العينات ذوات الحجم ن توزيع ذى الحدين دليلاه ن ، ح وسطه الحسابي ن ح وتباينه ن ح ك حيث ك = ١ – ح . ونعلم من البند (٦ – ٤) أنه إذا كان المتغير العشوائي ر = $\frac{س}{ن}$ يعبر عن النسبة المشاهدة في العينة فإن توزيع الإحصاءة :

$$ص = \frac{ر - ح}{\sqrt{\frac{ح ك}{ن}}}$$

يقترب من التوزيع المعتدل المعيارى بشرط أن تكون ن كبيرة وألا تكون ح أو ك قريبة من الصفر .

في هذه الحالة يمكن إثبات أن الفترة :

$$\left(ر - \sqrt{\frac{ر(١ - ر)}{ن}}\ ع_{\frac{\alpha}{٢}}\ ،\ ر + \sqrt{\frac{ر(١-ر)}{ن}}\ ع_{\frac{\alpha}{٢}}\right) \qquad (١٩)$$

هى فترة ثقة بدرجة $(1 - \alpha)$ للنسبة ح .

حيث مع$_{\frac{\alpha}{2}}$ هى القيمة الحرجة في التوزيع المعتدل المعيارى التي تحقق المعادلة

$$ل\ (مع_{\frac{\alpha}{2}} > ع > -مع_{\frac{\alpha}{2}}) = 1 - \alpha \qquad (20)$$

الشكل (٦ – ٥)

فمثلا حين $\alpha = 0,05$ فإن مع$_{\frac{\alpha}{2}}$ = ١,٩٦ راجع المثال (٤ – ٣).

وحين $\alpha = 0,01$ فإن مع$_{\frac{\alpha}{2}}$ = ٢,٥٨ راجع المثال (٤ – ٣) .

ويلاحظ أن أى عدد يقع في هذه الفترة يصلح لأن يؤخذ كتقدير للنسبة ح .

مثال (٦ – ١٥) :

أخذت عينة عشوائية حجمها ٤٠٠ شخص من مجتمع ما ووجد أن ٢٣ منهم مصابون بمرض البول السكرى . أوجد فترة ثقة بدرجة ٩٥٪ لنسبة المصابين بهذا المرض في هذا المجتمع .

الحل :

$$ر = \frac{23}{400} = 0,0575 \quad \text{نسبة الإصابة في العينة}$$

نأخذ هذه النسبة كتقدير للنسبة ح مع ملاحظة عشوائية العينة وكبر حجمها .

$$\text{تقدير الخطأ المعيارى} = \sqrt{\frac{ر(1-ر)}{ن}} = \sqrt{\frac{0,0575 \times 0,9425}{400}} = 0,0116$$

من (١٩) وبأخذ α = ٠,٠٥ نجد أن

الحد الأدنى للفترة = ٠,٠٥٧٥ − ٠,٠١١٦ × ١,٩٦ = ٠,٠٣٥

الحد الأعلى للفترة = ٠,٠٥٧٥ + ٠,٠١١٦ × ١,٩٦ = ٠,٠٨٠٢

إذن الفترة (٠,٠٣٥ ، ٠,٠٨٠) هي فترة ثقة بدرجة ٩٥٪ لنسبة المصابين بالبول السكرى في المجتمع .

(٦ – ٩ – ١) اختبار دلالة الفرق بين نسبتي عينتين مستقلتين :

نفرض أن لدينا مجتمعين نسبة وقوع حدث ما في أحدهما $ح_١$ ونسبة وقوعه في الآخر $ح_٢$. ونفرض أننا حصلنا على عينتين عشوائيتين مستقلتين من هذين المجتمعين ووجدنا أن نسبتي وقوع الحدث فيهما $ر_١$ ، $ر_٢$ على الترتيب . يمكن إثبات أن توزيع المعاينة للإحصاءة .

$$ص = \frac{(ر_١ - ر_٢) - (ح_١ - ح_٢)}{\sqrt{\frac{ح_١ ك_١}{ن_١} + \frac{ح_٢ ك_٢}{ن_٢}}} \qquad (٢١)$$

يقترب من التوزيع المعتدل المعيارى كلما كبرت كل من $ن_١$ ، $ن_٢$ ولم تكن أى من $ح_١$ ، $ك_١$ ، $ح_٢$ ، $ك_٢$ قريبة من الصفر .

وإذا فرضنا أن $ح_١ = ح_٢ = ح$ ، $ك = ١ - ح$ فإن هذه الإحصاءة تكتب .

$$ص = \frac{ر_١ - ر_٢}{\sqrt{ح ك \left(\frac{١}{ن_١} + \frac{١}{ن_٢}\right)}} \qquad (٢١)$$

وفى هذه الحالة تقدر ح من العينتين كالآتي :

$$ح = \frac{ر_١ ن_١ + ر_٢ ن_٢}{ن_١ + ن_٢}$$

وعلى ذلك يمكن استخدام الإحصاءة (٢١) للكشف عما إذا كان الفرق المشاهد بين $س_١$ ، $س_٢$ في العينتين هو فرق صغير لا دلالة له أم فرق يدل على وجود فرق حقيقي بين النسبتين $ح_١$ ، $ح_٢$ في المجتمعين .

مثال (٦ – ١٦) :

لمعرفة تأثير طُعم ما في الوقاية من وباء الكوليرا اختيرت عينتان عشوائيتان حجم الأولى ١٠٠٠ شخص وحجم الثانية ١٥٠٠ شخص وحقنت أفراد المجموعة الأولى بالطعم ولم يحقن أفراد المجموعة الثانية (مجموعة مراقبة) . وبعد فترة من الزمن ظهرت ١٠٠ حالة مرضية في المجموعة الأولى و ٥٠٠ حالة في المجموعة الثانية . اختبر ما إذا كان لهذا الطعم أثر في الوقاية من هذا المرض مستخدماً مستوى الدلالة ٠,٠١

الحل :

الفرض الصفرى $ف_٠$: $ح_١ = ح_٢$ (لا يوجد تأثير للطعم المعطى)

الفرض الآخر $ف_١$: $ح_١ < ح_٢$ (اختبار ذو جانب واحد) .

نسبة الإصابة في المجموعة الأولى $س_١ = \frac{١٠٠}{١٠٠٠} = ٠,١٠$

نسبة الإصابة في المجموعة الثانية $س_٢ = \frac{٥٠٠}{١٥٠٠} = ٠,٣٣$

من (٢١) : تقدير $ح = \frac{٠,١ \times ١٠٠٠ + ٠,٣٣ \times ١٥٠٠}{١٠٠٠ + ١٥٠٠} = ٠,٢٣٨$

تقدير الخطأ المعياري $= \sqrt{ح\, ك \left(\frac{١}{ن_١} + \frac{١}{ن_٢}\right)} = \sqrt{٠,٢٣٨ \times ٠,٧٦٢ \left(\frac{١}{١٠٠٠} + \frac{١}{١٥٠٠}\right)}$

$= ٠,٠١٧$

من (٢١) : $ص = \frac{٠,٣٣ - ٠,١٠}{٠,٠١٧} = ١٣,٥٢$

بما أن −١٣,٥٢ < −٢,٣٣ نرفض الفرض الصفرى عند المستوى ٠,٠١ ونستنتج أن للطعم تأثير في الوقاية من المرض .

(٦ – ٩ – ٢) فترات الثقة للفرق بين نسبتي مجتمعين :

يمكن إثبات أنه تحت الشروط سابقة الذكر تكون الفترة :

$$\left\{ (\hat{\text{ح}}_1 - \hat{\text{ح}}_2) - \text{خ. م.}\ \text{مع}_{\frac{\alpha}{2}} \ ،\ (\hat{\text{ح}}_1 - \hat{\text{ح}}_2) + \text{خ. م.}\ \text{مع}_{\frac{\alpha}{2}} \right\} \qquad (٢٢)$$

حيث $\text{خ. م.} = \sqrt{\bar{\text{ح}}\ \text{ل} \left(\frac{1}{\text{ن}_1} + \frac{1}{\text{ن}_2} \right)}$

هى فترة ثقة بدرجة $(1 - \alpha)$ للفرق $\text{ح}_1 - \text{ح}_2$ بين نسبتي المجتمعين . ففي المثال (٦ – ١٥) الأخير نجد أن :

الحد الأدني للفترة = (٠,٣٣ − ٠,١) − ٠,٠١٧ × ٢,٥٨ = ٠,١٨٦

الحد الأعلى للفترة = (٠,٣٣ − ٠,١) + ٠,٠١٧ × ٢,٥٨ = ٠,٢٧٤

أى أن الفترة (٠,١٨٦ ، ٠,٢٧٤) هى فترة ثقة بدرجة ٩٩٪ للفرق بين النسبتين في المجتمعين .

تمارين (٦ – ٣)

(١) في عينة عشوائية من ٤٠٠ شخص حقنوا بمصل ما تأثر ١٣٦ تأثيراً ضاراً . أوجد فترة ثقة بدرجة ٩٥٪ لنسبة الأشخاص الذين يتأثرون تأثيراً ضاراً من هذا المصل .

(٢) من ١٠٠ سمكة صيدت من بحيرة ما وجد أن ١٦ منها لا تصلح للأكل نتيجة لتلوث كيميائي في بيئة هذه البحيرة . أنشيء فترة ثقة بدرجة ٩٥٪ لاحتمال أنه إذا صيدت سمكة من هذه البحيرة تكون غير صالحة للأكل .

(٣) في تجربة عن تهجين فيران ذوات الشعر النحيف المتهدل وفيران ذوات الشعر المجعد وجد في ٣٢ ولادة بكل منها ٨ فيران أن عدد الولادات التي تحتوى بالضبط على س فأراً ذا شعر نحيف ومتهدل كالآتي :

س :	٠	١	٢	٣	٤	٥	٦	٧	٨
عدد الولادات :	٠	١	٢	٤	١٢	٦	٥	٢	٠

على فرض أن توزيع ذى الحدين يصلح نموذجاً في هذه التجربة أوجد فترة ثقة بدرجة ٩٩٪ لاحتمال الحدث « ولادة فأر ذى شعر نحيف ومتهدل » – راجع البند (٣ – ٣ – ١) .

(٤) أخذت مجموعتان عشوائيتان بكل منهما ٨٠ مريضاً . أعطى للمجموعة الأولى أقراص تحتوى على دواء ضد الحساسية وأعطى للمجموعة الثانية أقراص تمويه (لا تحتوى على أى دواء) ، فظهرت أعراض الحساسية في ٢٣ من المجموعة الأولى ، ٥١ من المجموعة الثانية . اختبر عند مستوى الدلالة ٠,٠١ ما إذا كانت نسبة ظهور الحساسية في المرضي الذين يتلقون أقراص الحساسية لا تختلف عنها في المرضي الذين يتلقون أقراص التمويه .

(٥) في عينة عشوائية من ٢٠٠ شخص لا يتناولون طعام الإفطار أفاد ٨٢ منهم أنهم يشعرون بتعب منتصف الصباح . وفي عينة عشوائية من ٤٠٠ شخص يتناولون طعام الإفطار أفاد ١١٦ منهم أنهم يشعرون بتعب منتصف الصباح . استخدم مستوى الدلالة ٠,٠١ لاختبار الفرض الصفرى أنه لا يوجد فرق بين المجتمعين ضد الفرض الآخر أن تعب منتصف الصباح أكثر تفشياً بين الأشخاص الذين لا يتناولون طعام الإفطار .

(٦ – ١٠) تحديد حجم العينة :

من الأمور التي تشغل الباحث عند تصميم تجربة لحل مشكلة ما تحديد عدد

وحدات العينة (العشوائية) اللازم لضمان أن تكون أحكامه عن المجتمع الذى يدرسه على درجة كافية من العمومية والدقة . وبطبيعة الحال كلما كانت العينة كبيرة كلما زادت الثقة في هذه الأحكام ، غير أن كبر حجم العينة يحتاج إلى الكثير من الجهد والوقت والتكاليف ، سواء في عملية المعاينة أو في قياس وتحليل البيانات ولذلك فإن كفاءة التصميم تتطلب إيجاد حد أعلى معقول لحجم العينة . وسنبحث هذا الموضوع في الحالات الثلاث الآتية .

(أولا) عند تقدير الوسط الحسابي لمجتمع :

نفرض أننا نتساءل عن حجم العينة المناسب لتقدير متوسط المجتمع μ من المتوسط $\bar{س}$ لعينة عشوائية مأخوذة منه . إن الإجابة عن هذا التساؤل لا تتأتى إلا إذا أجبنا عن السؤال العملى الآتى : « ما مقدار الخطأ الذى يمكن السماح به عند تقدير μ عن طريق $\bar{س}$ ؟ أى ما هو الحد الأعلى الذى يمكن التجاوز عنه لانحراف $\bar{س}$ عن القيمة الحقيقية μ » ؟ وهذا السؤال لا يجاب عنه إحصائيا وإنما هو من اختصاص الباحث التطبيقى وهو الذى يجيب عنه من واقع خبرته بميدان البحث . فإذا رأى الباحث أن الحد الأعلى للخطأ المسموح به هو عدد ما خ ، ورأى فى الوقت نفسه أن يعين درجة ثقة (٩٥٪ مثلا) فى عدم تخطى هذا الحد عند التطبيق ، فإن الحجم المناسب للعينة الذى يحقق الفرض المنشود ينتج حسب الأساس الآتى :

من البند (٦ – ٣) نعلم أنه إذا كان لدينا مجتمع معتدل وسطه الحسابي μ وانحرافه المعيارى σ فإن الإحصاءة $\bar{س}$ للعينات ذوات الحجم ن يكون لها توزيع معتدل وسطه الحسابى μ وانحرافه المعيارى $\frac{\sigma}{\sqrt{ن}}$ وتتحقق هذه النظرية أيضا (بالتقريب) حين لا يكون المجتمع معتدلا بشرط أن يكون حجم العينة كبيرا (ن $\geqslant$ ٣٠) .

وإذا كانت $\bar{س}$ هى الوسط الحسابي لعينة عشوائية ما فإن الفترة :

$$(\bar{س} - \frac{\sigma}{\sqrt{ن}} \text{مع}_{\frac{\alpha}{2}} \text{ ، } \bar{س} + \frac{\sigma}{\sqrt{ن}} \text{مع}_{\frac{\alpha}{2}})$$

تكون فترة ثقة بدرجة $1 - \alpha$ للمتوسط μ للمجتمع ، حيث مع $_{\frac{\alpha}{2}}$ هى قيمة المتغير المعتدل المعيارى التي تحقق المعادلة :

$$ل(\text{مع}_{\frac{\alpha}{2}} > ع > \text{مع}_{\frac{\alpha}{2}}) = 1 - \alpha$$

وبذلك تكون أكبر قيمة للانحراف $|\bar{س} - \mu|$ هى $\frac{\sigma}{\sqrt{ن}}$ مع $_{\frac{\alpha}{2}}$ ويسمى هذا العدد بحد الخطأ error bound ، بدرجة ثقة $1 - \alpha$

وإذا اخترنا أن يكون الخطأ المسموح به هو مقدار معين خ وأردنا أن نكون على ثقة بدرجة $(1 - \alpha)$ بألا يتعدى الخطأ الذى تقع فيه القيمة خ فإن حجم العينة المطلوب يجب أن يحقق المعادلة $\frac{\sigma}{\sqrt{ن}}$ مع $_{\frac{\alpha}{2}}$ = خ . وبحل هذه المعادلة في ن نجد أن :

$$ن = \left(\frac{\text{مع}_{\frac{\alpha}{2}} \sigma}{خ}\right)^2 \qquad (23)$$

وهذه هى القيمة المطلوبة لحجم العينة ن الذى يكفي لتحقيق الغرض المطلوب .

كما يمكن أن نأخذ العدد ن حيث

$$ن = \frac{\sigma^2}{\alpha خ^2} \qquad (24)$$

كحد أعلى لحجم العينة . وتنتج هذه المعادلة من متباينة تشبيشف الشهيرة التي لا تتطلب توزيعاً أو شروطاً معينة ، إلا أنها غالباً ما تعطى قيمة أكبر مما ينبغى لحجم العينة .

ويلاحظ أن إيجاد قيمة ن من أى من المعادلتين (٢٣) أو (٢٤) يتطلب معرفة ولو تقريبية بالانحراف المعيارى σ للمجتمع . وحين تكون قيمة σ مجهولة تماماً فلا مفر من تقديرها من عينة عشوائية استطلاعية كبيرة لا يقل حجمها عن ٣٠ .

مثال (٦ – ١٧) :

في عينة عشوائية حجمها ٣٦ وجد أن الوسط الحسابي والانحراف المعيارى هما ٢,٦ و٠,٣ على الترتيب . ما حجم العينة اللازم لإعطائنا ثقة بدرجة ٩٥٪ بألا يزيد الخطأ في تقدير الوسط الحسابي μ للمجتمع عن ٠,٠٦ ؟

الحل :

لدينا خ = ٠,٠٦ ، ١ – α = ٠,٩٥ ومنها α = ٠,٠٥ ،

مع$_{\frac{\alpha}{2}}$ = ١,٩٦

نأخذ الانحراف المعيارى ٠,٣ الناتج من العينة الاستطلاعية كتقدير للانحراف المعيارى للمجتمع . ثم نحسب ن من الصيغة (٢٣) كالآتي :

$$ن = \left(\frac{١,٩٦ \times ٠,٣}{٠,٠٦}\right)^{٢} = ٩٦,٠٤$$

إذن الحجم المطلوب للعينة هو ٩٧ أى أنه يكفينا أن نوجد الوسط الحسابي $\bar{س}$ لعينة عشوائية من هذا الحجم لكى نكون واثقين بدرجة ٩٥٪ أن اختلاف $\bar{س}$ عن μ لا يزيد عن ٠,٠٦ .

نلاحظ أنه بأخذ $ن = ٩٧$ فإن $\frac{ع}{\sqrt{ن}} \times ١,٩٦ = \frac{١,٩٦ \times ٠,٣}{\sqrt{٩٧}} = ٠,٠٥٩٧$

وهذا أصغر من ٠,٠٦ .

(ثانيا) عند تقدير الفرق بين متوسطى مجتمعين معتدلين :

بنفس الطريقة يمكن إيجاد حد أعلى لحجم العينة اللازم لتقدير الفرق $\mu_1 - \mu_2$ بين متوسطى مجتمعين معتدلين بحيث نكون على ثقة بدرجة $١ - \alpha$ بألا يزيد المقدار $\bar{س}_1 - \bar{س}_2$ المشاهد فى عينتين عن المقدار $\mu_1 - \mu_2$ للمجتمعين ، عن قيمة معينة خ مفروضة مسبقا ، وما علينا إلا استخدام الخطأ المعيارى لتوزيع المعاينة للإحصاءة $\bar{س}_1 - \bar{س}_2$ وهو $\sqrt{\frac{\sigma_1^2}{ن_1} + \frac{\sigma_2^2}{ن_2}}$ بدلا من الخطأ المعيارى $\frac{\sigma}{\sqrt{ن}}$ لتوزيع المعاينة للإحصاءة $\bar{س}$. وإذا كانت العينتان متساويتى الحجم : $ن_1 = ن_2 = ن$ وكان المجتمعان متساويين فى التباين : $\sigma_1^2 = \sigma_2^2 = \sigma^2$ فإن ذلك الخطأ المعيارى يأخذ الصورة $\frac{\sqrt{٢\sigma^2}}{\sqrt{ن}}$ ويكون الحد الأعلى المطلوب لحجم كل من العينتين هو

$$ن = \left(\frac{\sqrt{٢}\,\sigma\, مع_{\frac{\alpha}{٢}}}{خ}\right)^٢ \qquad (٢٥)$$

ويلاحظ أن هذه الصيغة ضعف الصيغة (٢٣) .

مثال (٦ — ١٨) :

فى تجربة تهدف إلى اختبار دواء تخسيس وتحكم فى زيادة الوزن عند النساء فى البدء باستخدام الدواء على إناث القطط المنزلية ، فاختيرت عينة عشوائية من

٦٠ قطة قسمت عشوائيا إلى مجموعتين متكافئتين فى الوزن بكل منهما ٣٠ قطة ، مجموعة تجريبية وأخرى ضابطة ، أعطى الدواء إلى المجموعة التجريبية فقط ووضعت المجموعتان تحت نفس الظروف . وبعد ٦ أسابيع وباستخدام وحدة قياس معينة وجد أن متوسط النقص فى وزن المجموعة التجريبية ١٠ وحدات ومتوسط النقص فى وزن المجموعة الضابطة ١٣,٣ وحدة . على فرض أن تباين كل من المجموعتين ٤٠ أوجد حدا أعلى لحجم كل من العينتين الذى يعطينا ثقة بدرجة ٩٥٪ بألا يزيد الفرق المشاهد فى متوسطى العينتين عن الفرق الحقيقى بين متوسطى المجموعتين عن ٣ وحدات .

الحل :

لدينا $\sigma^{٢} = ٤٠$ ، خ $= \pm ٣$ ، $\alpha = ٠,٠٥$ وإذن $مع_{\frac{\alpha}{٢}} = ١,٩٦$

من الصيغة (٢٥) : ن $= \frac{٤٠ \times ٢}{٩} (١,٩٦)^{٢} = ٣٤,١٤٨ \sim ٣٥$

(ثالثا) عند تقدير نسبة وقوع حدث في مجتمع :

نفرض أن نسبة وقوع حدث معين في مجتمع هو مقدار ثابت مجهول ح ونفرض أننا نرغب في معرفة حجم العينة المناسب لتقدير هذه النسبة عن طريق النسبة حَ التي تظهر في عينة عشوائية ، بحيث نكون على ثقة بدرجة ١ $- \alpha$ ألا يزيد الخطأ الناشيء عن هذا التقدير عن مقدار معين خ .

من البند (٦ — ٤) نعلم أنه إذا كان حجم العينة كبيراً (ن $\geqslant$ ٣٠) فإن توزيع المعاينة للنسبة حَ لوقوع هذا الحدث في العينات ذوات الحجم ن يقترب من توزيع معتدل وسطه الحسابي ح وانحرافه المعيارى $\sqrt{\frac{ح ك}{ن}}$ حيث ك = ١ — ح . في

هذه الحالة يكون حد الخطأ أى أكبر قيمة للمقدار $|حَ - ح|$ هو $\sqrt{\frac{ح ك}{ن}} . مع_{\frac{\alpha}{٢}}$ ،

يلاحظ أن هذا الحد هو نفس الحد الذى وجدناه في (أولا) بوضع $\sqrt{\frac{ح ك}{ن}}$ بدلا من $\frac{\sigma}{\sqrt{ن}}$

بحل المعادلة $\sqrt{\frac{ح ك}{ن}} \; مع_{\frac{\alpha}{2}} = خ$ نحصل على

$$ن = ح ك \left(\frac{مع_{\frac{\alpha}{2}}}{ح}\right)^{٢} \qquad (٢٦)$$

وهذه هى القيمة المطلوبة لحجم العينة . غير أن هذا الحل غير قابل للاستخدام لأنه يشتمل على البارامتر ح وهو الذى نبحث عن تقديره . ولكن نظراً لأن أكبر قيمة لحاصل الضرب ح ك هى $\frac{١}{٤}$ أى أن ح ك $\leq \frac{١}{٤}$ دائماً ،

[لأن ٤ ح ك − ١ = ٤ ح (١ − ح) − ١ = − (٤ ح^{٢} − ٤ ح + ١)
= − (٢ ح − ١)^{٢} $\leq$ ٠]

فإن الصيغة (٢٦) يمكن أن تكتب كالآتي :

$$ن = \frac{١}{٤} \left(\frac{مع_{\frac{\alpha}{2}}}{خ}\right)^{٢} \qquad (٢٧)$$

واختيار الحد الأقصي لحجم العينة من هذه المعادلة يؤدى الغرض المنشود بغير أى معرفة مسبقة لقيمة البارامتر ح .

أما إذا كان لدينا معلومات تفيد بأن هذا البارامتر يساوى بالتقريب قيمة معينة ح مثلا فإن ن يمكن إيجادها من المعادلة .

$$ن = \bar{ح} \bar{ك} \left(\frac{مع_{\frac{\alpha}{2}}}{خ}\right)^{٢} \qquad (٢٨)$$

وهذه القيمة تقل عن تلك التي تعطيها الصيغة (٢٧) لأنها مبنية على معلومات عن القيمة المحتملة للدليل ح .

مثال (٦ – ١٩) :

في عملية مسح صحي عن طريق العينة ، يراد تقدير النسبة ح للأشخاص ضعاف البصر . كم شخصاً ينبغي اختبارهم إذا كان المسئولون يريدون أن يتأكدوا بدرجة ٩٨٪ أن الخطأ في التقدير يقع في المدى ± ٠,٠٥ .

(أ) في حالة عدم وجود أي معرفة لقيمة ح

(ب) إذا كان المعروف أن ح = ٠,٣ بالتقريب ؟

الحل :

لدينا خ = ± ٠,٠٥ ، ١ − α = ٠,٩٨ إذن α = ٠,٠٢ ، $\frac{\alpha}{٢}$ = ٠,٠١

القيمة هي قيمة المتغير المعتدل المعيارى أ التى تحقق المعادلة

ل (أ > ع > − أ) = ٠,٩٨

ل (أ > ع > ٠) = ٠,٩٨ ÷ ٢ = ٠,٤٩

من جدول المساحات أسفل المنحني المعتدل نجد أن أ = ٢,٣٣

(أ) من (٢٧) : ى = $\frac{١}{٤}$ $\left(\frac{٢,٣٣}{٠,٠٥}\right)^{٢}$ = ٥٤٢,٨٩

إذن الحجم المطلوب للعينة هو ٥٤٣ .

(ب) من (٢٨) : ى = ٠,٣ × ٠,٧ $\left(\frac{٢,٣٣}{٠,٠٥}\right)^{٢}$ = ٤٥٦,٠٢٧٦

إذن الحجم المطلوب للعينة هو ٤٥٦ .

تمارين (٦ – ٤)

(١) يريد باحث تقدير متوسط محتوى الفسفات في وحدة حجم من ماء إحدى البحيرات والمعروف من دراسات سابقة أن الانحراف المعيارى لهذا المحتوى ذو قيمة

تكاد تكون ثابتة عند σ = ٤ . ما عدد عينات الماء التي ينبغى للباحث تحليلها لكى يثق بدرجة ٩٠٪ أن الخطأ في التقدير لا يتعدى ٠,٨ ؟

(٢) أ – أراد مهندس تقدير الوسط الحسابي للمدة التي تجف فيها خلطة معينة من الأسمنت تستخدم في إصلاح الطرق . وقد جرب هذه الخلطة في ١٠٠ بقعة ووجد أن الوسط الحسابي والانحراف المعيارى لمدة الجفاف هما ٣٢ دقيقة و ٤ دقائق على الترتيب . استخدم هذه المعلومات لإيجاد فترة ثقة بدرجة ٩٩٪ لمتوسط مدة جفاف هذه الخلطة .

ب – أراد المهندس أن يحدد حجم العينة (أى عدد البقع التي يجرب فيها الخلطة) لكى يكون متأكداً بدرجة ٩٥٪ أن متوسط مدة الجفاف المحسوب من عينة بهذا الحجم لا يختلف عن المتوسط الحقيقي لهذه المدة إلا بدقيقة واحدة على الأكثر . أوجد هذا الحجم علماً بأن الخبرة السابقة تشير إلى أن الانحراف المعيارى لمدة الجفاف هو ٥ دقائق بالتقريب .

(٣) في أحد مراكز القلب والصدر يراد تقدير المعدل ح لوقوع حالات ضيق التنفس بين الذكور من متوسطى العمر ممن كانوا يدخنون أكثر من علبتين من السجائر في اليوم خلال خمس سنوات سابقة . ما حجم العينة التي نختاره بحيث نكون على ثقة بدرجة ٩٥٪ أن الخطأ في تقدير النسبة ح لا يزيد عن ٠,٠٣ ، علماً بأن المعروف أن قيمة خ الحقيقية قريبة من ٠,١٥ ؟

(٦ – ١١) مراقبة الانتاج OUALITY CONTROL

من تطبيقات نظرية العينات فى الصناعة ذلك التطبيق المسمى بمراقبة جودة الإنتاج ومعاينات القبول ، وهو يتعلق بإحدى المشكلات التى تظهر فى المصانع التى تنتج سلعا على نطاق واسع . ويتمثل هذا التطبيق فى استخدام فكرة فترات الثقة فى التفتيش على جودة السلع المنتجة أثناء عملية الإنتاج .

ومن المعروف أن الوحدات المنتجة فى أى مصنع لا يمكن أن تكون جميعها متشابهة تماما مهما تقدمت التكنولوجية الصناعية بل توجد دائما انحرافات صغيرة عن المواصفات الموضوعة للسلعة تنشأ عن عدد كبير من العوامل الصغيرة التى لا يمكن التحكم فيها وتعتبر عوامل عشوائية ، ولا مفر للمصنع من أن يسمح بالتجاوز عن هذه الانحرافات طالما كانت فى حدود معقولة يقبلها المنتج والمستهلك ، أما إذا خرجت الانحرافات عن هذه الحدود فإن المصنع يشتبه فى وجود خلل ما إما فى أجزاء آلة المصنع أو فى سلوك العمال أو فى الإدارة أو أية مصادر أخرى للأخطاء ، وعليه حينئذ أن يتحرى عن أسباب هذا الخلل ويعمل على تلافيه . وليس من المعقول أن يفحص المصنع كل وحدة ينتجها والمتبع أن يجرى الآتى .

نفرض مثلا أن مصنعا ينتج يوميا عشرات الألوف من الأنابيب الاسطوانية ذات مواصفات معينة منها أن طول الاسطوانة ٦ سنتيمترات مثلا . ليطمئن المصنع على توفر هذه الصفة يضع اختبارا إحصائيا ليختبر به الفرض أن الوحدات المنتجة تتوفر فيها الخاصة المطلوبة ، أى ليختبر الفرض الصفرى μ = ٦ سم ضد الفرض $\mu \neq$ ٦ سم ويجرى هذا الاختبار فى فترات منتظمة من الزمن ، مثلا كل نصف ساعة أو كل ساعة ، وفى كل مرة يأخذ عينة عادة من ٣ إلى ١٠ وحدات ويقيس متوسط الطول $\bar{س}$ فيها و يطبق الاختبار . فإذا أدى الاختبار إلى قبول الفرض الصفرى فإن هذا يعنى أن الإنتاج يسير سيرا طبيعيا ، أما إذا أدى إلى رفض الفرض الصفرى فإن هذا يعنى أن انحرافات أطوال الوحدات المنتجة عن الطول المطلوب هى انحرافات جوهرية وينبغى حينئذ التحرى عن أسباب هذه الانحرافات . وهذا ما يسمى بمراقبة الإنتاج .

وتوضع قاعدة الاختبار على هيئة فترة ثقة مركزها القيمة μ المطلوبة (أى أن حدى الثقة يكونان على بعدين متساويين من μ) . وإذا افترضنا أن توزيع الأطوال فى الوحدات المنتجة **معتدل** متوسطه μ وانحرافه المعيارى σ فإن توزيع المعاينة

للمتوسطات الحسابية للعينات التى من الحجم ن يكون معتدلا متوسطه μ وانحرافه المعيارى $\frac{\sigma}{\sqrt{\text{ن}}}$ – راجع البند (٦ – ٣ – أولا) – وتكون الفترة

$$\left(\bar{\text{س}} - \text{٢,٥٨} \times \frac{\sigma}{\sqrt{\text{ن}}} \text{ ، } \bar{\text{س}} + \text{٢,٥٨} \times \frac{\sigma}{\sqrt{\text{ن}}}\right)$$

هى فترة ثقة بدرجة ٩٩٪ للمتوسط μ . أى أنه فى ٩٩٪ من العينات التى نأخذها تقع μ فى هذه الفترة ، أى تتحقق المتباينة الآتية :

$$\bar{\text{س}} - \text{٢,٥٨} \times \frac{\sigma}{\sqrt{\text{ن}}} < \mu < \bar{\text{س}} + \text{٢,٥٨} \times \frac{\sigma}{\sqrt{\text{ن}}}$$

ومن هذه المتباينة المزدوجة ينتج أن

$$\mu - \text{٢,٥٨} \times \frac{\sigma}{\sqrt{\text{ن}}} < \bar{\text{س}} < \mu + \text{٢,٥٨} \times \frac{\sigma}{\sqrt{\text{ن}}} \qquad (\text{٢٩})$$

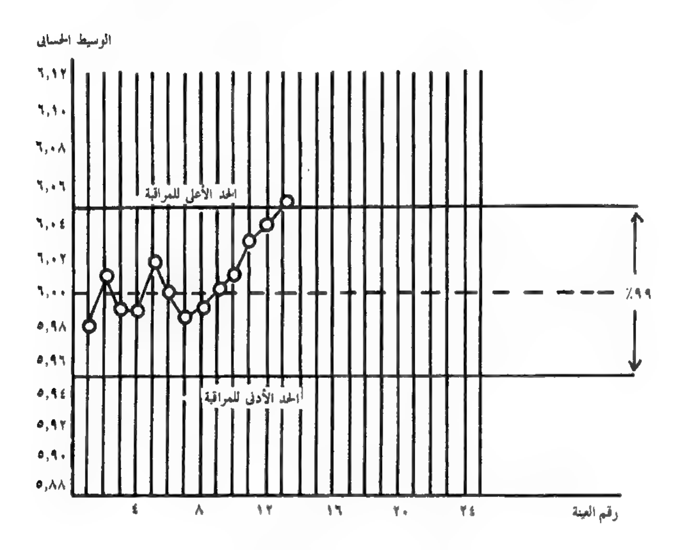
ويسمى الطرف الأيسر من هذه المتباينة **بحد المراقبة الأدنى** ، ويسمي الطرف الأيمن منها **بحد المراقبة الأعلى** lower and upper control limits وعند المراقبة إذا أخذت عينة ووجد أن متوسطها $\bar{\text{س}}$ يقع بين هذين الحدين يقبل الفرض الصفرى ويستمر الإنتاج دون تعديل ، أما إذا ساوت $\bar{\text{س}}$ أحد الحدين أو تعدته فيرفض الفرض الصفرى عند مستوى الدلالة ٠,٠١ وينبغى حينئذ اتخاذ ما يلزم (وربما إيقاف الآلة) لتحرى أسباب الخلل وإصلاحه .

وفى المعتاد تستخدم فترة ثقة بدرجة ٩٩٪ أما إذا أردنا استخدام فترة ثقة بدرجة ٩٥٪ فإننا نضع العدد ١,٩٦ بدلا من العدد ٢,٥٨ فى المتباينة (٢٩) – راجع المثال (٤ – ٣) – أما قيمة σ فيمكن تقديرها من عينة عشوائية كبيرة تؤخذ من الإنتاج عندما يكون فى حالته الطبيعية .

ولتسهيل عملية المراقبة ترصد قيمة المتوسط $\bar{س}$ لكل عينة مختبرة بالترتيب فى خريطة تسمى **بخريطة المراقبة** control chart . فإذا فرضنا أن μ = ٦سم ،

σ = ٠,٠٦سم ، ن = ١٠ فان حدى المراقبة يكونان ٦ ± $\frac{٠,٠٦}{\sqrt{١٠}}$ × ٢,٥٨

أى بالتقريب ٥,٩٥ ، ٦,٠٥ من السنتيمترات وتكون خريطة المراقبة كالآتى :



الشكل (٦-٦) : خريطة المراقبة للوسط الحسابى

وتبين هذه الخريطة الخطين الممثلين للحدين الأدنى والأعلى للمراقبة ، كما تمثل النقط متوسطات ١٣ عينة (حجم كل منها ١٠) بترتيب اختيارها وهى ٥,٩٨١ ، ٦,٠١١ ، ٥,٩٩١ ، ٥,٩٩٢ ، ٦,٠٢٢ ، ٦,٠٠٣ ، ٥,٩٨٥ ، ٥,٩٩٢ ، ٦,٠٠٤ ، ٦,٠١٣ ، ٦,٠٣٢ ، ٦,٠٤٣ ، ٦,٠٥٤ ، ويبدو من قيمة المتوسط الأخير أن الإنتاج فى حاجة إلى تعديل عند الوصول إلى العينة ١٣ .
هذا ويمكن بنفس الطريقة متابعة التغير فى تباين الإنتاج أو فى أى بارامتر آخر .

تمارين (٦ – ٥)

أخذت عشر عينات حجم كل منها ٤ وحدات فى فترات منتظمة ووجد أن سمك الوحدات كما فى الجدول الآتى . مثل متوسطات هذه العينات على خريطة مراقبة على فرض أن المجتمع معتدل متوسطه ٥ وانحرافه المعيارى ١,٥٥ .

الزمن	٨/–	٨/٣٠	٩/–	٩/٣٠	١٠/–	١٠/٣٠	١١/–	١١/٣٠	١٢/–	١٢/٣٠
سمك الوحدة	٣	٣	٥	٧	٧	٤	٥	٦	٥	٥
	٤	٦	٢	٥	٣	٤	٦	٤	٥	٢
	٨	٦	٥	٤	٦	٣	٤	٦	٦	٥
	٤	٨	٦	٤	٥	٦	٦	٤	٤	٣

الفصل السابع

حساسية اختبارات الفروض

SENSITIVITY OF TESTS OF HYPOTHESES

لا يوجد قرار إحصائى منزه عن الخطأ ، فالقرارات الإحصائية هى دائما قرارات احتمالية بمعنى أنه لا مفر من وجود احتمال للخطأ فى أى قرار نصدره عن مجتمع عن طريق عينة . ولما كانت هذه القرارات مؤسسة على ما نجريه من اختبارات للفروض وتزيد ثقتنا فيها بزيادة حساسية هذه الاختبارات ، وجب علينا أن ندرس كيف نزيد من هذه الحساسية أى من قدرة الاختبارات على تمكيننا من اتخاذ القرار السليم الذى لا يشوبه إلا قدر ضئيل من الخطأ . ويتأتى ذلك عن طريق التحكم ما أمكن فى احتمالات الأخطاء التى تنجم حتما عند استخدام هذه الاختبارات . ومن الطبيعى إذن أن نبدأ بتقديم هذه الأخطاء توطئة لدراسة كيفية التقليل منها ما استطعنا إلى ذلك سبيلا .

(٧ – ١) نوعا الأخطاء الإحصائية :

نعلم حتى الآن أننا حين نكون بصدد اتخاذ قرار برفض أو قبول فرض صفرى ف. ضد فرض آخر ف١ عند مستوى معين α من الدلالة ، نقوم أولا بتجزىء فضاء العينة إلى منطقتين منفصلتين ومتكاملتين م ، م̄ نسمى إحداهما م بمنطقة الرفض أو بالمنطقة الحرجة ونسمى الأخرى م̄ بمنطقة القبول . وإذا وقعت قيمة مشاهدة من إحصاءة الاختبار فى المنطقة م رفضنا الفرض الصفرى عند المستوى

α لصالح الفرض الآخر ، أما إذا وقعت القيمة المشاهدة فى المنطقة م َ فإننا نقبل الفرض الصفرى .

ونظرا لأننا لا نعرف مسبقا ما إذا كان الفرض الصفرى صحيحا أو زائفا فإن القرار الذى نتخذه يكون على إحدى الحالات الآتية :

(١) أن يكون الفرض الصفرى صحيحا ويكون قرارنا هو رفض هذا الفرض .
(٢) أن يكون الفرض الصفرى صحيحا ويكون قرارنا هو قبول هذا الفرض .
(٣) أن يكون الفرض الصفرى زائفا ويكون قرارنا هو رفض هذا الفرض .
(٤) أن يكون الفرض الصفرى زائفا ويكون قرارنا هو قبول هذا الفرض .

ومن الواضح أن قرارنا يكون صائبا فى الحالتين (٢) و(٣) ويكون خاطئا فى الحالتين (١) و(٤) . يقال للخطأ الناشىء عن القرار (١) إنه خطأ من النوع الأول ، كما يقال للخطأ الناشىء عن القرار (٤) إنه خطأ من النوع الثانى . ويعرّف هذان النوعان من الخطأ كالآتى :

الخطأ من النوع الأول : TYPE I ERROR

هو ذلك الخطأ الذى ينشأ حين نتخذ قرارا برفض الفرض الصفرى بينما يكون هذا الفرض صحيحا فى الواقع . ويرمز لاحتمال هذا الخطأ بالرمز α .

الخطأ من النوع الثانى : TYPE II ERROR

هو ذلك الخطأ الذى ينشأ حين نتخذ قرارا بقبول الفرض الصفرى بينما يكون هذا الفرض زائفا فى الواقع . ويرمز لاحتمال هذا الخطأ بالرمز β .

وقد اصطلح على تسمية المقدار ه = ١ − β بقوة الاختبار power of the test وهو يعبر عن احتمال تجنب الخطأ من النوع الثانى . ومن الواضح أن قوة الاختبار تزيد كلما نقص الاحتمال β للخطأ من النوع الثانى والعكس بالعكس .

نلخص المعانى السابقة فى الجدول (٧ - ١) الآتى :

الجدول (٧ - ١)

حالات رفض أو قبول الفرض الصفرى

القرار	الحالة المجهولة للفرض الصفرى ف. صحيح	 ف. زائف
رفض ف.	خطأ I (باحتمال α)	صواب (باحتمال ١ - β)
قبول ف.	صواب (باحتمال ١-α)	خطأ II (باحتمال β)

يلاحظ أن كلا من الاحتمالين α ، β هو احتمال شرطى ونكتب :

α = ل (رفض ف. | ف. صحيح) ، β = ل (قبول ف. | ف. زائف) .

فى الأمثلة التى تناولناها فى الفصل السابق كان اهتمامنا منصبا على الخطأ من النوع الأول وحرصنا على أن يكون الاحتمال α لهذا الخطأ عددا صغيرا يسمح لنا بالتجاوز عن هذا الخطأ ، وسمينا هذا العدد مستوى الدلالة واتخذناه كأحد أسس قاعدة اختبار الفروض ، خاصة فيما يتعلق بفصل فضاء العينة إلى منطقتى رفض وقبول الفرض الصفرى . وسنهتم الآن بالخطأ من النوع الثانى ، إذ ينبغى عند تصميم التجارب وبناء اختبارات الفروض أن نعمل على أن يكون كل من الاحتمالين α ، β صغيرا على قدر الإمكان .

غير أنه نظرا لأن تصغير أحد هذين الاحتمالين يؤدى إلى كبر الآخر كما سنرى فى البند (٧ - ٢ - ٢) ، يكون من العبث البحث عن طريقة عامة تضمن صغر كل من هذين الاحتمالين معا ونكون حينئذ أمام مشكلة يجب أن نجد لها حلا .

والطريقة المعتادة لتناول هذه المشكلة تبدأ بالتحكم فى احتمال الخطأ من النوع الأول (وهو الخطأ الأكثر خطورة) وذلك بوضع حد أعلى للاحتمال α فنختار لهذا الحد قيمة صغيرة مثل ٠,٠٥ أو ٠,٠١ ثم نحسب على أساسها كلا من منطقتى قبول ورفض الفرض الصفرى من واقع ما لدينا من بيانات ومن معرفتنا بتوزيع إحصاءة الاختبار . بعد ذلك نقوم بحساب الاحتمال β ، فإذا كان هذا الاحتمال صغيرا تكون المشكلة قد حلت تلقائيا ، أما إذا كان كبيرا بدرجة لا نستطيع معها المجازفة به وجب علينا أن نعمل على تخفيض هذا الاحتمال . وهذا ما سنتناوله فى البند (٧ — ٥) بعد تقديم طريقة حساب هذا الاحتمال والعوامل المؤثرة عليه .

طريقة إيجاد احتمال الخطأ من النوع الثانى :

نفرض أننا اخترنا مستوى الدلالة α وحددنا عند هذا المستوى كلا من منطقة الرفض م ومنطقة القبول مَ للفرض الصفرى مع ملاحظة أن بارامتر إحصاءة الاختبار يتحدد هنا على أساس التسليم بصحة هذا الفرض . بعد ذلك نوجد احتمال وقوع قيم المتغير فى منطقة القبول مَ على أساس أن الفرض الصفرى زائف وأن الفرض الآخر هو الصحيح فيكون هذا الاحتمال هو احتمال قبول الفرض الصفرى عندما يكون زائفا أى هو الاحتمال β للخطأ من النوع الثانى مع ملاحظة اختلاف بارامتر إحصاءة الاختبار . وعلى هذا فحساب الاحتمال β يكون على خطوتين هما :

(أ) تحديد منطقة القبول على أساس صحة الفرض الصفرى ،

(ب) إيجاد احتمال وقوع المتغير فى هذه المنطقة على أساس أن الفرض الآخر هو الصحيح .

سنوضح هذا الأسلوب لحساب قيمة الاحتمال β فى عدة حالات نبدأها فى البند (٧ — ٢) بحالة الاختبار ذى الجانب الواحد لفرض عن متوسط مجتمع معتدل

مع بيان العوامل المؤثرة على قوة الاختبار وكيفية زيادة هذه القوة ، ثم نطبق ذلك كله على ثلاث حالات أخرى نقدمها فى البنود الثلاثة الأخيرة .

(٧ – ٢) حساب قيمة β فى اختبار فرض عن متوسط معتدل – حالة الاختبار ذى الجانب الواحد .

مثال (٧ – ١) :

اعتبر مجتمعا معتدلا تباينه σ^2 = ٤٩ ووسطه الحسابى μ غير معروف . بين كيف تستخدم الوسط الحسابى $\bar{س}$ لعينة عشوائية من الحجم ن = ٢٥ لاختبار الفرض الصفرى μ = ٧٥ ضد الفرض الآخر μ > ٧٥ عند مستوى الدلالة ٥٪ ثم أوجد قوة الاختبار عندما μ = ٧٦ .

الحل :

نظرا لأن المجتمع معتدل فإن المتغير $\bar{س}$ الذى يعبر عن الوسط الحسابى للعينات من الحجم ن يكون ذا توزيع معتدل وسطه الحسابى μ وانحرافه المعيارى $\frac{\sigma}{\sqrt{ن}}$ – راجع البند (٦ – ٣) – وبالتالى يكون للإحصاءة

$$ع = \frac{\bar{س} - \mu}{\sigma/\sqrt{ن}}$$

توزيع معتدل معيارى : مع (٠ ، ١) .

لدينا σ^2 = ٤٩ ، ن = ٢٥ ، α = ٠,٠٥ وإذن $\frac{\sigma}{\sqrt{ن}} = \frac{٧}{٥}$ = ١,٤

الفرض الصفرى ف. : μ = ٧٥

الفرض الآخر ف١ : μ > ٧٥

نظرا لأن الاختبار ذو جانب واحد هو الجانب الأيمن فإن منطقة الرفض م هى

تلك المنطقة التي يأخذ فيها الوسط الحسابى $\bar{س}$ لعينة من الحجم ٢٥ قيمة أكبر من العدد ا حيث :

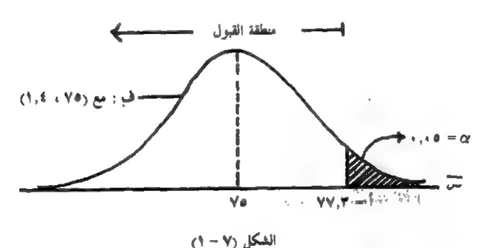
$$ل\ (\bar{س} > ا) = ٠,٠٥ \quad \text{تحت الفرض الصفرى} \ \mu = ٧٥$$

$$\text{ومنها}\ ل\ \left(\frac{\bar{س} - ٧٥}{١,٤} > \frac{ا - ٧٥}{١,٤}\right) = ٠,٠٥$$

$$\text{أى}\ ل\ \left(ع > \frac{ا - ٧٥}{١,٤}\right) = ٠,٠٥$$

من جدول المساحات أسفل المنحنى المعتدل المعيارى نجد أن

$$\frac{ا - ٧٥}{١,٤} = ١,٦٤ \qquad \text{ومنها}\ ا = ٧٧,٢٩٦ = ٧٧,٣ \ \text{تقريبا}$$



الشكل (٧ − ١)

منطقتا القبول والرفض فى اختبار ذى جانب واحد

وإذن المنطقة التى نرفض فيها الفرض الصفرى μ = ٧٥ حين يكون هذا الفرض صحيحا وعند المستوى α = ٠,٠٥ هى تلك المنطقة التى تأخذ فيها الأوساط الحسابية للعينات ذوات الحجم ٢٥ قيما تزيد عن ٧٧,٣ واحتمال وقوع المتغير

فى هذه المنطقة هو ٥٪،وتعبر عن هذا الاحتمال مساحة الجزء المظلل بالشكل (٧ — ١) . وعلى ذلك تتحدد قاعدة الاختبار كالآتى :

« إذا وقع الوسط الحسابى س̄ لعينة عشوائية من الحجم ٢٥ فى المنطقة م = { س̄ : س̄ > ٧٧,٣ } نرفض الفرض الصفرى ف. عند مستوى الدلالة ٥٪ وإلا نقبل ف. » .

أما منطقة قبول الفرض الصفرى فهى بالطبع المنطقة س̄ ≤ ٧٧,٣ المكملة لمنطقة الرفض . أى أننا نقبل الفرض الصفرى إذا كان الوسط الحسابى لعينة عشوائية من الحجم ٢٥ مأخوذة من التوزيع الممثل لهذا الفرض يساوى أو يقل عن ٧٧,٣ ونسبة هذه العينات هى ٩٥٪ من العينات التى تؤخذ من هذا التوزيع .

ولكن ماذا لو كان الفرض الصفرى زائفا والفرض الآخر هو الصحيح ؟ نفرض مثلا أن القيمة الصحيحة هى μ = ٧٦ (وهذه قيمة تحقق الفرض الآخر μ > ٧٥) .

فى هذه الحالة يكون للمتغير س̄ توزيع معتدل : مع (٧٦ ، ١,٤) ، وينتج ما يلى :

β = احتمال قبول الفرض الصفرى عندما يكون زائفا والفرض الآخر μ = ٧٦ هو الصحيح

= احتمال وقوع قيم المتغير س̄ فى منطقة القبول مَ حين يكون لهذا المتغير توزيع معتدل مع (٧٦ ، ١,٤)

= ل (س̄ ≤ ٧٧,٣)　　　　حيث س̄ : مع (٧٦ ، ١,٤)

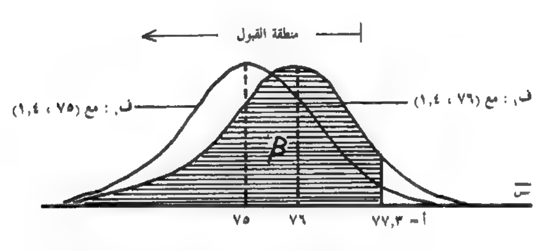
= ل ($\frac{\text{س̄} - ٧٦}{١,٤}$ ≤ $\frac{٧٧,٣ - ٧٦}{١,٤}$) = ل (ع ≤ ٠,٩٣)

من جدول المساحات أسفل المنحنى المعتدل المعيارى نجد أن

β = ٠,٨٢٣٨ = ٠,٨٢ تقريبا

وهذا هو احتمال الوقوع فى الخطأ من النوع الثانى . أما قوة الاختبار عندما تأخذ μ القيمة ٧٦ فهى ه = ١ − β = ٠,١٨ .

لتوضيح ذلك هندسيا نرسم كما فى الشكل (٧ − ٢) توزيع المتغير $\bar{س}$ فى حالتين ، أولاهما عندما يكون الفرض الصفرى ف صحيحا ويكون التوزيع معتدلا متوسطه ٧٥ وثانيهما عندما يكون الفرض الآخر ف١ صحيحا ويكون التوزيع معتدلا متوسطه ٧٦ .



الشكل (٧ − ٢)

التوزيع الممثل للفرض الصفرى والتوزيع الممثل للفرض الآخر

(مساحة الجزء المظلل تعبر عن الاحتمال β فى اختبار ذى جانب واحد)

من هذا الشكل يتبين أن بعض العينات التى تنتمى إلى توزيع ف١ تكون متوسطاتها واقعة فى منطقة القبول لتوزيع ف. ونسبة هذه العينات هى نسبة الجزء من توزيع ف١ الذى يشترك مع توزيع ف. فى منطقة القبول ، وهى تعطى بالاحتمال ل (س ≥ ٧٧,٣) محسوبا من توزيع ف١ وهذا الاحتمال هو بالضبط احتمال قبول الفرض الصفرى عندما يكون الفرض الآخر هو الصحيح ، أى هو الاحتمال β للخطأ من النوع الثانى .

(يلاحظ أننا لا نستطيع حساب قيمة β إلا إذا حددت قيمة معينة مثل μ =٧٦ للبارامتر μ تحقق الفرض الآخر(وهو μ > ٧٥٫٠) وذلك لكي يتحدد التوزيع الممثل للفرض الآخر تحديدا تاما) .

(٧ – ٢ – ١) دالة القوة POWER FUNCTION

فى هذا المثال وجدنا أن قوة الاختبار عندما نفترض أن μ = ٧٦ هى ٠,١٨ وتختلف هذه القوة بحسب قيمة μ فإذا افترضنا أن μ = ٧٧ نجد بنفس المنطق السابق أن

β = ل (سَ ≤ ٧٧,٣) حيث سَ : مع (٧٧ ، ١,٤)

= ل ((سَ – ٧٧) / ١,٤ ≤ (٧٧,٣ – ٧٧) / ١,٤) = ل (ع ≤ ٠,٢١)

= ٠,٥٨ وبالتالى يكون قوة الاختبار ﻫ = ١ – ٠,٥٨ = ٠,٤٢

وبالمثل : بأخذ μ = ٧٨ نجد أن β = ٠,٣١ و ﻫ = ٠,٦٩

وبأخذ μ = ٧٩ نجد أن β = ٠,١١ و ﻫ = ٠,٨٩

وبأخذ μ = ٨٠ نجد أن β = ٠,٠٣ و ﻫ = ٠,٩٧ ... الخ

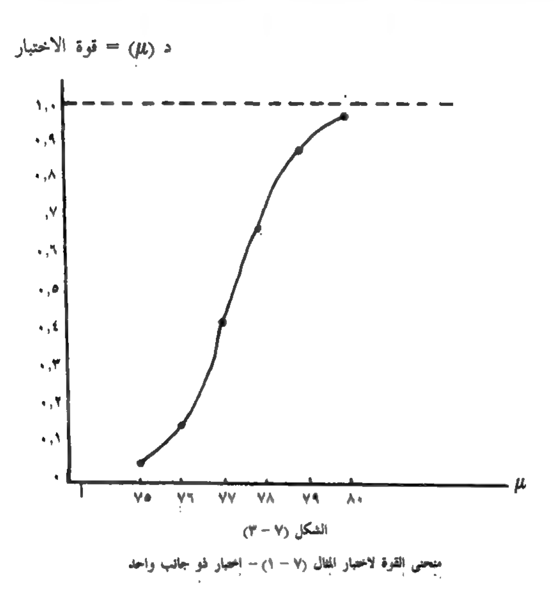
ومن هذا نرى أن قوة الاختبار تتوقف على البارامتر μ أى هى دالة د فى μ تأخذ الصيغة الآتية :

دالة قوة الاختبار = د (μ) = ﻫ = ١ – β = ١ – ل (سَ ≤ أ)

= ل (سَ > أ) حيث سَ : مع (μ ، ١,٤)

وحيث أ هى القيمة الحرجة التى تفصل بين منطقتى القبول والرفض . (قيمة الدالة د عند قيمة معينة μ تسمى قوة الاختبار عند القيمة μ) .

وتمثل دالة القوة بيانيا كما فى الشكل (٧ - ٣) الذى يوضح أنها دالة تزايدية ، تزيد قيمتها كلما بعدت قيمة μ التى يحددها الفرض الآخر عن قيمة μ التى يحددها الفرض الصفرى .



الشكل (٧ - ٣)

منحنى القوة لاختبار المثال (٧ - ١) - اختبار ذو جانب واحد

فى هذا المثال وجدنا أن احتمال الخطأ من النوع الثانى β = ٠,٨٢ ومن الواضح أن هذا الاحتمال هو احتمال كبير لهذا الخطأ لا ينبغى أن نسمح به حين نتخذ قرارا

بشأن رفض أو قبول الفرض الصفرى لأنه يجعلنا نشك فى حساسية الاختبار ، بمعنى أنه إذا كانت α = ٠,٠٥ ، σ^2 = ٤٩ وأخذنا عينة من الحجم n = ٢٥ فإن هذه العينة لا تكون قادرة على التمييز بين الفرضين بدرجة كافية من الثقة ، إذ بالرغم من أن ٩٥٪ من العينات المأخوذة من التوزيع الذى يفترض صحة الفرض الصفرى (μ = ٧٥) تقع فى منطقة القبول ، إلا أن ٨٢٪ من العينات المأخوذة من التوزيع الذى يفترض صحة الفرض الآخر ((μ = ٧٦) تقع أيضا فى المنطقة ذاتها . وهذا التداخل الكبير هو الذى نعنيه بقولنا إن الاختبار ذو قوة ضعيفة أو أنه اختبار غير حساس .

فى مثل هذه الحالة يجب أن ندخل تعديلا فى تصميم التجربة التى تمدنا بالبيانات التى نتخذ قرارنا على أساسها لتلافى الوقوع فى خطأ كبير من النوع الثانى ولإعطاء الاختبار قوة كافية للتمييز بين مختلف الفروض . وفى بحثنا عن التعديل اللازم لتحقيق هذا الغرض نبدأ بتدارس العوامل التى تؤثر فى هذا الخطأ .

(٧ – ٢ – ٢) العوامل المؤثرة فى الخطأ من النوع الثانى :

بالتأمل فى المثال السابق يتبين لنا أن الاحتمال β للخطأ من النوع الثانى وقوة الاختبار ف يتوقفان على القيم الآتية :

(١) القيمة التى تختار للاحتمال α للخطأ من النوع الأول :

ذلك لأن هذه القيمة هى التى تحدد القيمة الحرجة أ (تساوى ٧٧,٣ فى هذا المثال) التى تفصل بين منطقتى الرفض والقبول . وكلما صغرت قيمة α كلما صغرت منطقة الرفض وأزيحت أ إلى اليمين (فى هذا المثال) واتسعت منطقة القبول وبالتالى زاد احتمال هذه المنطقة تحت الفرض الآخر . ومن هنا يمكننا تقبل الحقيقة الآتية :

كلما صَغُر الاحتمال α للخطأ من النوع الأول كلما كبر الاحتمال β للخطأ من النوع الثانى وصغرت قوة الاختبار .

وهذه الحقيقة تعطينا سببا وجيها يبرر التقليد المتبع باختيار α = ٠,٠٥ أو α = ٠,٠١ . لأن اختيارنا لقيم تقل عن ذلك يؤثر تأثيرا سيئا فى الاحتمال β الذى يمكن التحكم فيه بتكبير حجم العينة كما سيتبين بعد .

(٢) قيمة البارامتر μ

بالتأمل فى قيم β أو ق التى حسبناها فى البند (٧ – ٢ – ١) نجد أن هذه القيم تتوقف على بعد القيمة μ_1 التى يحددها الفرض الآخر عن القيمة μ_0 التى يحددها الفرض الصفرى . ومن الناحية الهندسية إذا كانت μ_1 ، μ_0 قريبتين من بعضهما أى كان متوسطا توزيعى ف١ ، ف٠ قريبين من بعضهما فإن التداخل بين هذين التوزيعين فى منطقة القبول يكون كبيرا وهذا يؤدى إلى كبر الاحتمال β وصغر قوة الاختبار . أما إذا كانت μ_1 بعيدة عن μ_0 فإن هذا التداخل يكون صغيرا ويؤدى إلى صغر الاحتمال β وكبر قوة الاختبار . وتتضح هذه الحقيقة أيضا عند التأمل فى منحنى دالة القوة المبين بالشكل (٧ – ٣) . ومن هنا يمكننا تقبل الحقيقة الآتية :

كلما زاد الفرق بين القيمة التى يحددها الفرض الصفرى للبارامتر المختبر والقيمة التى يحددها الفرض الآخر ، كلما صَغُر الاحتمال β وزادت قوة الاختبار

(٣) الخطأ المعيارى لإحصاءة الاختبار :

فى المثال (٧ – ١) كان الخطأ المعيارى لإحصاءة الاختبار هو $\frac{\sigma}{\sqrt{n}}$ = ١,٤ ووجدنا أن قيمة العدد أ الذى يفصل بين منطقتى الرفض والقبول هى ٧٧,٣ . إذا

أجرينا تعديلا فى هذا المثال بحيث يصبح الخطأ المعيارى أقل من ١,٤ نجد أن ا تصغر وتزاح النقطة الممثلة لها على توزيع الإحصاءة $\bar{س}$ إلى اليسار وبالتالى تصغر منطقة القبول ويقل الاحتمال β . فمثلا إذا أخذنا $\frac{\sigma}{\sqrt{ن}}$ = ١,١ فإن

ل ($\bar{س}$ > ا) = ٠,٠٥ ومنها ل (ع > $\frac{ا - ٧٥}{١,١}$) = ٠,٠٥

∴ $\frac{ا - ٧٥}{١,١}$ = ١,٦٤ ∴ ا = ٧٦,٨٠٤ وهذا العدد أصغر من ٧٧,٣ ومن

هنا يمكننا تقبل الحقيقة الآتية :

كلما صَغُر الخطأ المعيارى لإحصاءة الاختبار
كلما صغر الاحتمال β وزادت قوة الاختبار .

ومن الواضح أن قيمة الكسر $\frac{\sigma}{\sqrt{ن}}$ تتوقف على قيمتين هما الإنحراف المعيارى σ للمجتمع وحجم العينة ن ، ويصغر هذا الكسر (أى يصغر الخطأ المعيارى) إذا صغرت قيمة σ فقط أو كبرت قيمة ن فقط أو صغرت σ وكبرت ن فى الوقت نفسه .

(٧ – ٢ – ٣) كيفية زيادة قوة الاختبار :

فى الفقرة السابقة وجدنا أن قوة الاختبار تتوقف على أربع قيم هى α ، μ ، σ ، ن . ولما كانت القيمتان α ، μ تتحددان سلفا بحسب خبرة الباحث وطبيعة المشكلة التى يتناولها ، لا يبقى لدينا من الناحية الإحصائية لزيادة قوة الاختبار إلا الاعتماد على تصغير الخطأ المعيارى $\frac{\sigma}{\sqrt{ن}}$ وذلك بتصغير σ أو تكبير ن .

ونظرا لأننا نقدر عادة تباين المجتمع σ^2 من تباين العينة فإن زيادة قوة الاختبار تقتضى أن نحرص على ألا يكون هذا التقدير أكبر مما ينبغى وهذا لا يتأتى إلا

بالتحكم الجيد فى ظروف التجريب واستبعاد تأثير أية عوامل خارجية تؤثر فى المشاهدات وتسهم فى زيادة تباينها .

أما زيادة حجم العينة فهو **العامل الرئيسى** الذى نعتمد عليه فى زيادة قوة الاختبار ، وهذه أهم نتيجة نخرج بها من هذا الفصل وتتلخص فى الحقيقة الآتية :

إذا تساوت جميع الظروف ، كلما كبر حجم العينة
كلما صغر الاحتمال β وزادت قوة الاختبار .

الحد الأمثل لحجم العينة :

على أن كفاءة التجريب تستدعى ألا يكون حجم العينة أكبر مما ينبغى تحسبا لما تتطلبه هذه العملية من جهد ووقت وتكاليف ، ومن المناسب إذن وضع حد أعلى لحجم العينة يحقق الهدف المنشود من زيادة قوة الاختبار دون تحمل أعباء لا ضرورة لها . غير أنه لا توجد قاعدة عامة لتحديد الحجم المناسب للعينة ، إلا أننا نستطيع ذلك فى بعض الحالات الخاصة ومنها الحالات التى يتناولها هذا الفصل .

ففى حالة اختبار فرض عن متوسط مجتمع معتدل ، نفرض أننا حددنا مسبقا قيمة α وقيمتى الفرض الصفرى والفرض الآخر . إذا أردنا أن نضمن أن يأخذ احتمال الخطأ من النوع الثانى قيمة محددة β يمكن إثبات أن الحد الأعلى لحجم العينة الذى يحقق هذا الضمان ينتج من حل المعادلة الآتية :

$$\text{خ . م .} = \frac{\text{ف}}{\text{ع}_1 - \text{ع}_2} \qquad (1)$$

حيث خ. م. = الخطأ المعيارى لإحصاءة الاختبار

، ف = الفرق بين القيمة التى يحددها الفرض الصفرى والقيمة التى يحددها الفرض الآخر

أما ع١ ، ع٢ فتحددان بحل المتباينتين الآتيتين :

$$ل (ع > ع_١) = \begin{cases} \alpha & \text{إذا كان الاختبار ذا جانب واحد} \\ \frac{\alpha}{٢} & \text{إذا كان الاختبار ذا جانبين} \end{cases}$$

$$، ل (ع \leqslant ع_٢) = \beta$$

حيث ع هو المتغير المعتدل المعيارى : مع (٠ ، ١) . أى أن ع١ ، ع٢ توجدان من جدول المساحات أسفل المنحنى المعتدل المعيارى بمجرد التعويض عن قيمتى α ، β . والمعادلة (١) هى معادلة عامة فى حالة اختبار فرض عن متوسط مجتمع معتدل أو اختبار الفرق بين متوسطى مجتمعين معتدلين ، مع بعض الاختلافات فى حساب القيم التى تتركب منها هذه المعادلة كما سنرى بعد .

مثال (٧ — ٢) :

اعتبر المثال (٧ — ١) حيث $\sigma^٢ = ٤٩$ ، $\alpha = ٠,٠٥$ نريد اختبار الفرض $\mu = ٧٥$ ضد الفرض $\mu > ٧٥$. وإذا أخذنا $\mu_١ = ٧٨$ فأوجد الحد الأعلى لحجم العينة الذى يضمن أن تكون $\beta = ٠,٠٨$.

الحل :

$$خ . م . = \frac{\sigma}{\sqrt{ن}} = \frac{٧}{\sqrt{ن}}$$

$$ف = \mu_١ - \mu = ٧٨ - ٧٥ = ٣$$

من جدول المساحات أسفل المنحنى المعتدل المعيارى نجد أن :

ل (ع > ع١) = α = ٠,٠٥ تعطى ع١ = ١,٦٤ (اختبار ذو جانب واحد)

، ل (ع $\leqslant$ ع٢) = β = ٠,٠٨ تعطى ع٢ = — ١,٤١

بالتعويض فى المعادلة (١) ينتج أن :

$$\frac{٧}{\sqrt{ن}} = \frac{٣}{١,٦٤ - (-١,٤١)} = \frac{٣}{٣,٠٥}$$

$$\therefore \quad ن = \left(\frac{٧ \times ٣,٠٥}{٣}\right)^{٢} = ٥٠,٦٤٦٩$$

أى أننا إذا أعدنا التجربة بأخذ عينة عشوائية من الحجم ٥١ (بدلا من الحجم ٢٥) فإننا نضمن أن يكون احتمال الخطأ من النوع الثانى β = ٠,٠٨ (تحقق من ذلك بأخذ ن = ٥١ وإثبات أن القيمة الحرجة أ = ٧٦,٦٠٧ والاحتمال β = ٠,٠٧٧٨) . يلاحظ أننا استطعنا أن نضمن احتمالا صغيرا هو ٠,٠٨ للخطأ من النوع الثانى (بدلا من الاحتمال β = ٠,٣١ الذى وجدناه فى عينة بالحجم ٢٥) ، إلا أن ذلك كان على حساب زيادة حجم العينة إلى الضعف تقريبا فى هذا المثال .

(٧ – ٣) حساب قيمة β فى اختبار فرض عن متوسط مجتمع معتدل – حالة الاختبار ذى الجانبين .

مثال (٧ – ٣) :

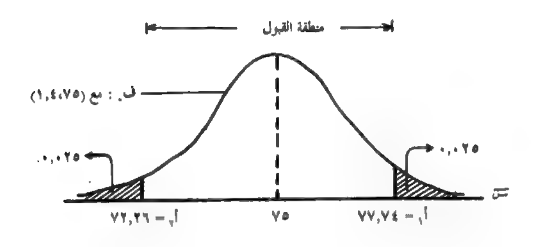
بأخذ بيانات المثال (٧ – ١) بيّن كيف تستخدم الوسط الحسابى لعينة لاختبار الفرض الصفرى ف. : μ = ٧٥ ضد الفرض الآخر $\mu \neq$ ٧٥ عند مستوى الدلالة ٠,٠٥ . حدد قوة الاختبار عند μ = ٧٦ وارسم دالة القوة لهذا الاختبار .

الحل :

نظرا لأن المجتمع معتدل يكون للمتغير $\bar{س}$ توزيع معتدل : مع (μ ، $\frac{\sigma}{\sqrt{ن}}$) وبالتالى يكون للاحصاءة

$$ع = \frac{\bar{س} - \mu}{\sigma/\sqrt{ن}}$$ توزيع معتدل معيارى : مع (٠ ، ١) .

ونظرا لأن الاختبار ذو جانبين فإن منطقة رفض الفرض الصفرى تتألف من جزءين متساويين فى جانبى التوزيع . لتكن $أ_١$ هى القيمة الحرجة التى تحد منطقة الرفض اليمنى من اليسار ، ولتكن $أ_٢$ هى القيمة الحرجة التى تحد منطقة الرفض اليسرى من اليمين . انظر الشكل (٧ - ٤) .



الشكل (٧ - ٤)

منطقتا الرفض والقبول فى اختبار ذى جانبين

لإيجاد قيمتى $أ_١$ ، $أ_٢$ نستخدم جدول المساحات أسفل المنحنى المعتدل المعيارى كالآتى :

$$ل\,(\bar{س} > أ_١) = \frac{٠,٠٥}{٢} = ٠,٠٢٥ \quad،\quad حيث\ \bar{س} : مع\,(٧٥ \,،\, ١,٤)$$

$$\therefore ل\,(\frac{\bar{س} - ٧٥}{١,٤} > \frac{أ_١ - ٧٥}{١,٤}) = ٠,٠٢٥ \quad أى\ ل\,(ع > \frac{أ_١ - ٧٥}{١,٤}) = ٠,٠٢٥$$

$$\therefore \frac{أ_١ - ٧٥}{١,٤} = ١,٩٦ \qquad ومنها\ أ_١ = ٧٧,٧٤٤$$

بالمثل ، ل $(\bar{س} < أ_٢) = \frac{٠,٠٥}{٢} = ٠,٠٢٥$ حيث $\bar{س}$: مع (٧٥ ، ١,٤)

$\therefore$ ل $(ع < \frac{أ_٢ - ٧٥}{١,٤}) = ٠,٠٢٥$

$\therefore \frac{أ_٢ - ٧٥}{١,٤} = - ١,٩٦$ ومنها $أ_٢ = ٧٢,٢٥٦$

وإذن تتحدد منطقة قبول الفرض الصفرى حين يفترض صحته بالفترة (٧٢,٢٦ ، ٧٧,٧٤) وتكون قاعدة الاختبار كالآتى :

« إذا وقع الوسط الحسابى لعينة عشوائية من الحجم ٢٥ خارج المنطقة م َ = $\{\bar{س} : ٧٧,٧٤ \geqslant \bar{س} > ٧٢,٢٦\}$ نرفض الفرض الصفرى ف عند مستوى الدلالة ٠,٠٥ وإلا نقبل ف »

نحسب الآن الاحتمال β عند $\mu = ٧٦$ وفقا للمنطق السابق .

β = احتمال وقوع قيم المتغير $\bar{س}$ فى منطقة القبول (٧٢,٢٦ ، ٧٧,٧٤) تحت الفرض الآخر $\mu = ٧٦$

= ل $(٧٧,٧٤ > \bar{س} > ٧٢,٢٦)$ حيث س : مع (٧٦ ، ١,٤)

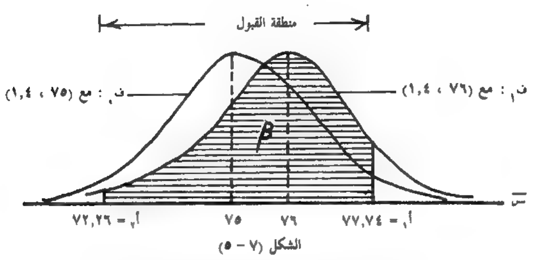
= ل $(\frac{٧٦ - ٧٧,٧٤}{١,٤} \leqslant \frac{٧٦ - \bar{س}}{١,٤} \leqslant \frac{٧٦ - ٧٢,٢٦}{١,٤})$

= ل $(١,٢٤ \leqslant ع \leqslant - ٢,٦٧) = ٠,٨٩$ تقريبا

$\therefore$ قوة الاختبار عند $\mu = ٧٦$ هى و = ١ − ٠,٨٩ = ٠,١١ تقريبا

لتوضيح ذلك هندسيا نرسم كما فى الشكل (٧ − ٥) توزيع المتغير $\bar{س}$ تحت كل من الفرضين ف : $\mu = ٧٥$ وف : $\mu = ٧٦$. من هذا الشكل يلاحظ

أن الاحتمال β = ٠,٨٩ هو نسبة الجزء من توزيع ف$_{١}$ الذى يشترك فى منطقة القبول مع توزيع ف$_{٠}$ ، وهذه النسبة تعبر عنها مساحة الجزء المظلل من الشكل (٧ — ٥) .



الشكل (٧ — ٥)

التوزيع الممثل للفرض الصفرى والتوزيع الممثل للفرض الآخر
(مساحة الجزء المظلل تعبر عن الاحتمال β فى اختبار ذى جانبين)

فى هذه الحالة تأخذ دالة قوة الاختبار الصيغة الآتية :

د (μ) = ل ($\bar{س}$ ≥ أ$_{١}$) + ل ($\bar{س}$ ≤ أ$_{٢}$)

= ١ — ل (أ$_{١}$ > $\bar{س}$ > أ$_{٢}$)

وفى هذا المثال نجد أن :

د (μ) = ١ — ل (٧٧,٧٤ > $\bar{س}$ > ٧٢,٢٦)   حيث $\bar{س}$: مع (μ ، ١,٤)

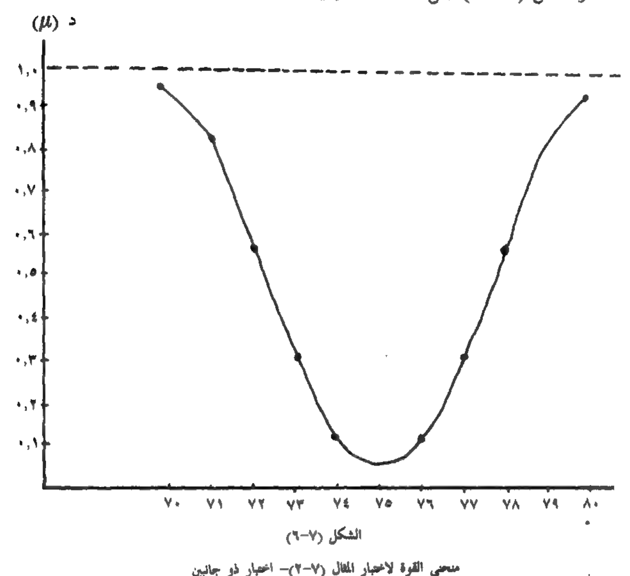
وقد وجدنا أن د (٧٦) = ١ — ٠,٨٩ = ٠,١١

وبالمثل نجد ما يلى :

د (٧٧) = ٠,٣٠ و د (٧٨) = ٠,٥٧ و د (٧٩) = ٠,٨٢ و د (٨٠) = ٠,٩٤ ... الخ . ومن التماثل نجد أن

د (٧٤) = ٠,١١ و د (٧٣) = ٠,٣٠ و د (٧٢) = ٠,٥٧ و د (٧١) = ٠,٨٢ و د (٧٠) = ٠,٩٤ ... الخ

والشكل (٧ – ٦) يمثل هذه الدالة بيانيا .



الشكل (٧–٦)

منحنى القوة لاختبار المثال (٧–٢)– اختبار ذو جانبين

إن الملاحظات والحقائق التى ذكرت فى البند (٧ – ٢ – ٢) عن العوامل المؤثرة فى قيمة β تنطبق هنا أيضا . فكلما صغُر الاحتمال α كلما صَغُر جزءا منطقة الرفض ، واتسعت منطقة القبول وهذا يؤدى إلى زيادة الاحتمال β وصغر قوة الاختبار . كذلك كلما بعدت القيمة μ التى يفرضها الفرض الآخر عن القيمة التى يحددها الفرض الصفرى كلما قل التداخل بين توزيعى ف. ، ف١ ، وصغرت

وزادت قوة الاختبار . وأخيرا كلما نقص الخطأ المعيارى سواء بتصغير الانحراف المعيارى σ للمجتمع أو بزيادة حجم العينة ، كلما صغرت β وزادت قوة الاختبار .

وجدير بالملاحظة هنا أنه إذا تساوت جميع الظروف فإن الاحتمال β للخطأ من النوع الثانى يكون أقل فى الاختبار ذى الجانب الواحد منه فى الاختبار ذى الجانبين ، أى أن الاختبار ذا الجانب الواحد يكون أقل تعرضا لهذا النوع من الخطأ .

وكما فى البند (٧ – ٢ – ٣) حين نتناول اختبارا ذا جانبين لفرض عن متوسط مجتمع معتدل ، إذا تحددت قيم α ، μ ، σ وأردنا أن نضمن أن يأخذ احتمال الخطأ من النوع الثانى قيمة محددة β فإن الحد الأعلى لحجم العينة الذى يوفر هذا الضمان هو ذلك الذى ينتج من حل نفس المعادلة (١) السابق تقديمها وهى :

$$\text{خ.م.} = \frac{\text{ف}}{\text{ع}_1 - \text{ع}_2}$$

والفرق فى استخدام هذه المعادلة بين الحالتين يحدث فقط فى حساب القيمة ع_1

مثال (٧ – ٤) :

فى المثال (٧ – ٣) حيث $\sigma^2 = ٤٩$ ، $\alpha = ٠,٠٥$ ، $\mu = ٧٥$ إذا أخذنا $\mu_1 = ٧٨$ فأوجد الحد الأعلى لحجم العينة الذى يضمن أن تكون $\beta = ٠,١٠$.

الحل :

$$\text{خ.م.} = \frac{\sigma}{\sqrt{\text{ن}}} = \frac{٧}{\sqrt{\text{ن}}}$$

$$\text{ف} = \mu_1 - \mu = ٧٨ - ٧٥ = ٣$$

ل (ع > ع$_1$) = $\frac{\alpha}{2}$ = ٠,٠٢٥ ومنها ع$_1$ = ١,٩٦ (اختبار ذو جانبين)

ل (ع ≥ ع$_2$) = β = ٠,١٠ ومنها ع$_2$ = − ١,٢٨

$$\therefore \quad \frac{٧}{\sqrt{ن}} = \frac{٣}{١,٩٦ - (-١,٢٨)} = \frac{٣}{٣,٢٤} = \frac{١}{١,٠٨}$$

ومنها ن = (٧ × ١,٠٨)^2 = ٥٧,١٥

أى أنه يكفى أن نأخذ عينة حجمها ٥٨ لنضمن أن تكون β = ٠,١٠ (تحقق من ذلك بأخذ ن = ٥٨ وإثبات أن القيمتين الحرجتين هما ا$_1$ = ٧٦,٨٠ ، ا$_2$ = ٧٣,١٩٩ وأن β = ٠,٠٩٨١)

(٧ − ٤) حساب قيمة β فى اختبار الفرق بين متوسطى مجتمعين معتدلين :

نستخدم نفس الأسلوب المقدم فى البندين السابقين مع مراعاة طريقة حساب الخطأ المعيارى التى تتطلبها هذه الحالة .

مثال (٧ − ٥) :

لتجربة أقراص لإنقاص الوزن عند النساء ، اختيرت ٤٠ من إناث القطط المنزلية وقسمت عشوائيا إلى مجموعتين بكل منها ٢٠ قطة ووضعت المجموعتان تحت نفس الظروف والنظام الغذائى فيما عدا أن الأقراص كانت تضاف إلى غذاء واحدة فقط من المجموعتين . وبعد ستة أسابيع وباستخدام مقياس معين حسب الوسطان الحسابيان $\bar{س}_1$ ، $\bar{س}_2$ للمجموعتين . إذا اعتبرنا أن المجموعتين مستقلتان ومأخوذتان من مجتمعين معتدلين تباين كل منها ٤٠ ومتوسطاهما μ_1 ، μ_2 :

(أولا) بين كيف نختبر الفرض الصفرى $\mu_1 = \mu_2$ ضد الفرض $\mu_1 \neq \mu_2$ عن مستوى الدلالة ٠,٠٥ .

(ثانيا) أوجد قوة الاختبار عندما يفترض أن الفرق بين متوسطى المجتمعين يساوى ٣ وحدات .

(ثالثا) أوجد الحد الأدنى لحجم العينة الذى يضمن أن يكون الاحتمال β للخطأ من النوع الثانى يساوى ٠,١٥ .

الحل :

(أولا)

نظرا لأن المجتمعين معتدلان والعينتين مستقلتان فإن المتغير العشوائى $\bar{س}_١ - \bar{س}_٢$ يكون ذا توزيع معتدل وسطه الحسابى $\mu_١ - \mu_٢$ وتباينه يساوى

$$\frac{\sigma_١^٢}{ن_١} + \frac{\sigma_٢^٢}{ن_٢} = \frac{٢\sigma^٢}{ن} \quad \text{لأن } \sigma_١^٢ = \sigma_٢^٢ = \sigma^٢ \text{ و } ن_١ = ن_٢ = ن$$

وبالتالى يكون للإحصاءة

$$ع = \frac{(\bar{س}_١ - \bar{س}_٢) - (\mu_١ - \mu_٢)}{\sqrt{\frac{٢\sigma^٢}{ن}}}$$

توزيع معتدل معيارى : مع (٠ ، ١) .

لدينا $\sigma_١^٢ = \sigma_٢^٢ = ٤٠$ و $ن_١ = ن_٢ = ٢٠$

اذن $\sqrt{\frac{٢\sigma^٢}{ن}} = \sqrt{\frac{٢ \times ٤٠}{٢٠}} = ٢$ وهذا هو الخطأ المعيارى للإحصاءة

الفرض الصفرى ف. هو $\mu_1 = \mu_2$

الفرض الآخر ف، هو $\mu_1 \neq \mu_2$ (اختبار ذو جانبين) ، α = ٠,٠٥

نظرا لأن الاختبار ذو جانبين فإن منطقة رفض الفرض الصفرى تتألف من منطقتين عند ذيلى التوزيع يحدهما العددان $أ_1$ ، $أ_2$ بحيث

ل $(\bar{س}_1 - \bar{س}_2 < أ_1)$ = ٠,٠٢٥ و ل $(\bar{س}_1 - \bar{س}_2 > أ_2)$ = ٠,٠٢٥

على أن يحسب كلا الاحتمالين على أساس التسليم بصحة الفرض الصفرى $\mu_1 = \mu_2$ أى من الإحصاءة

$$ع_ه = \frac{(\bar{س}_1 - \bar{س}_2) - صفر}{٢} = \frac{\bar{س}_1 - \bar{س}_2}{٢}$$

التى تتبع التوزيع المعتدل المعيارى : مع (٠ ، ١) .

لدينا ل $(\bar{س}_1 - \bar{س}_2 < أ_1)$ = ل $\left(\frac{\bar{س}_1 - \bar{س}_2}{٢} < \frac{أ_1}{٢}\right)$ = ل $\left(ع < \frac{أ_1}{٢}\right)$

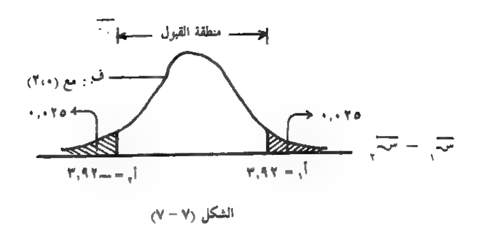
= ٠,٠٢٥

من جدول المساحات أسفل المنحنى المعتدل نجد أن

$\frac{أ_1}{٢}$ = ١,٩٦ ومنها $أ_1$ = ٣,٩٢

من التماثل نجد أن :

$\frac{أ_2}{٢}$ = −١,٩٦ ومنها $أ_2$ = −٣,٩٢



الشكل (٧ – ٧)

وإذن تتألف المنطقة التي ترفض فيها الفرض الصفرى ف : $\mu_1 = \mu_2$ عند المستوى ٠,٠٥ من اتحاد المنطقتين ($\bar{س}_1 - \bar{س}_2$) >٣,٩٢ و($\bar{س}_1 - \bar{س}_2$)< –٣,٩٢ أى من المنطقة التي يأخذ فيها الفرق بين الوسطين الحسابيين للعينات من الحجم ٢٠ قيما تزيد عن ٣,٩٢ أو قيما تقل عن – ٣,٩٢ واحتمال هاتين المنطقتين معا تعبر عنه مساحتى الجزءين المظللين بالشكل (٧ – ٧) ، وتكون منطقة القبول هى المنطقة المكملة لهاتين المنطقتين أى المنطقة التي تحددها الفترة (– ٣,٩٢ ، ٣,٩٢) وتتحدد قاعدة الاختبار كالآتى :

« إذا وقع الفرق بين متوسطى عينتين مستقلتين من الحجم ٢٠ خارج المنطقة م′ = {$\bar{س}_1 - \bar{س}_2$: ٣,٩٢ ⩾ $\bar{س}_1 - \bar{س}_2$ ⩾ – ٣,٩٢} نرفض الفرض الصفرى ف٠ عند مستوى الدلالة ٠,٠٥ وإلا نقبل ف٠ . »

(ثانيا)

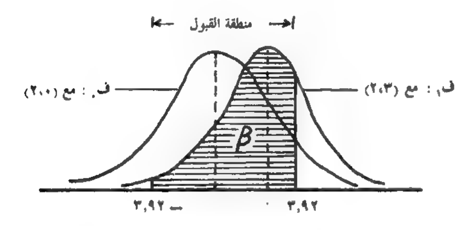
لحساب قيمة β نحسب احتمال وقوع قيم المتغير $\bar{س}_1 - \bar{س}_2$ فى منطقة القبول على أساس أن الفرض الصفرى خاطىء والفرض الآخر $\mu_1 - \mu_2$ = ٣ هو الصحيح .

β = ل (٣,٩٢ ⩾ $\bar{س}_1 - \bar{س}_2$ ⩾ – ٣,٩٢) حيث $\bar{س}_1 - \bar{س}_2$: مع (٣ ، ٢)

$$= ل \left(\frac{٣,٩٢ - ٣}{٢} \geqslant \frac{\bar{س}_١ - \bar{س}_٢ - ٣}{٢} \geqslant \frac{- ٣,٩٢ - ٣}{٢} \right)$$

= ل (٠,٤٦ ⩾ ع ⩾ -٣,٤٦) = ٠,٦٨ من جدول المساحات

وإذن قوة الاختبار عندما $\mu - \mu_٢$ = ٣ هى ه = ٠,٣٢ - انظر الشكل (٧ - ٨) .



الشكل (٧ - ٨)

التوزيع الممثل للفرض الصفرى والتوزيع الممثل للفرض الآخر

(مساحة الجزء المظلل تعبر عن الاحتمال β للخطأ من النوع الثانى)

(ثالثا)

لإيجاد الحد الأعلى لحجم العينة الذى يضمن أن تكون β = ٠,١٥ نستخدم نفس المعادلة (١) السابق تقديمها وهى :

$$خ . م . = \frac{ف}{ع_١ - ع_٢}$$

ويستلزم الأمر هنا أن تكون العينتان مستقلتين ومن نفس الحجم .

لدينا خ . م . = $\sqrt{\frac{2\sigma^2}{n}}$ = $\sqrt{\frac{80}{n}}$

، ف = ٣ − ٠ = ٣

ل (ع > ع$_1$) = $\frac{\alpha}{2}$ = ٠,٠٢٥ لأن الاختبار ذو جانبين

وإذن ع$_1$ = ١,٩٦

ل (ع ⩽ ع$_2$) = β = ٠,١٥ وإذن ع$_2$ = ١,٠٤

∴ $\sqrt{\frac{80}{n}}$ = $\frac{3}{1.96 - (-1.04)}$ = $\frac{3}{3}$ = ١

∴ ن = ٨٠

أى أنه يكفى أخذ عينتين مستقلتين حجم كل منهما ٨٠ لكى نضمن أن تكون β = ٠,١٥ (تحقق من ذلك بأخذ ن = ٨٠ وإثبات أن ا$_1$ = ١,٩٦ و ا$_2$ = −١,٩٦ وأن β = ٠,١٤٩٢)

(٧ − ٥) حساب قيمة β فى اختبار النسبة :

فى البنود الثلاثة السابقة كنا نتناول الأوساط الحسابية لعينات من مجتمعات معتدلة أو معتدلة تقريبا . على أن المنطق الذى استخدمناه فى حساب الاحتمال β للخطأ من النوع الثانى وحساب قوة الاختبار ينطبق على أى مقاييس أخرى . وفى المثال الآتى نتناول نسبة وقوع حدث ما فى مجتمع ما .

مثال (٧ − ٦) :

بينت الخبرة أن معدل الشفاء من مرض معين بواسطة علاج قياسى ٦٠٪ ابتكر علاج جديد يظن أنه أفضل من العلاج القياسى . بين كيف تختبر عند مستوى

الدلالة ٠,٠٥ ما إذا كان معدل الشفاء بالعلاج الجديد أعلى منه بالعلاج القياسى ، وذلك باستخدام عينة من ١٥ مريضا بهذا المرض . حدد قوة الاختبار عندما يفترض أن معدل الشفاء بالعلاج الجديد ٧٠٪ .

الحل :

إن جودة العلاج تقاس بقيمة المتغير س الذى يعبر عن عدد المرضى الذين شفوا فى عينة من الحجم ١٥ . وإذا اعتبرنا أن العينة عشوائية ذات وحدات مستقلة فإن المتغير س يكون له توزيع ذى الحدين دليلاه ن ، ح حيث ن = ١٥ ، ح بارامتر مجهول يعبر عن احتمال الشفاء لأى مريض . راجع البند (٣ — ٣) .

ولبحث أفضلية العلاج الجديد ، علينا أن نقارن بين الفرضين الآتيين :

الفرض الصفرى ف. : ح = ٠,٦ (لا يوجد فرق فى معدل الشفاء بين نوعى العلاج)

الفرض الآخر ف١ : ح > ٠,٦ (اختبار ذو جانب واحد)

إذا كان الفرض الصفرى صحيحا ، يكون للمتغير س توزيع ذو حدين دليلاه ن = ١٥ ، ح = ٠,٦ ويمكننا حينئذ إيجاد توزيع احتمال هذا المتغير بالحساب المعتاد (أى من دالة الكتلة) أو باستخدام الجدول (٣) فى ذيل هذا الكتاب مع أخذ ن = ١٥ ، ح = ٠,٦ فنجد التوزيع الذى ننقله فى الجدول (٧ — ٢) الآتى .

نظرا لأن الاختبار ذو جانب واحد هو الجانب الأيمن فإن منطقة رفض الفرض الصفرى هى المنطقة التى يأخذ فيها المتغير س قيما تزيد عن العدد ا حيث ل (س > ا) لا تزيد عن α = ٠,٠٥ ولإيجاد القيمة الحرجة ا التى تحد التوزيع من اليمين نجرب بضعة قيم مستعينين بالجدول (٧ — ٢) كالآتى :

ل (س > ١٤) = ٠

ل (س > ١٣) = ٠,٠٠٥

ل (س > ١٢) = ٠,٠٠٥ + ٠,٠٢٢ = ٠,٠٢٧
ل (س > ١١) = ٠,٠٢٧ + ٠,٠٦٣ = ٠,٠٩٠

وهذا الاحتمال يزيد عن ٠,٠٥

إذن القيمة الحرجة أ١ = ١٢ وتتحدد قاعدة الاختبار كالآتى :

« إذا كان عدد المرضى الذين شفوا فى عينة من ١٥ مريضا يزيد عن ١٢ نرفض الفرض الصفرى ف. أن ح = ٠,٦ عند مستوى الدلالة ٠,٠٥ وإلا نقبل ف. » .

الجدول (٧ -٢)

توزيع الاحتمال لمتغير ذى حدين : حد (١٥ ، ٠,٦)

س عدد حالات الشفاء	ل احتمال هذا العدد	س عدد حالات الشفاء	ل احتمال هذا العدد
٠	ــــ	٨	٠,١٧٧
١	ــــ	٩	٠,٢٠٧
٢	ــــ	١٠	٠,١٨٦
٣	٠,٠٠٢	١١	٠,١٢٧
٤	٠,٠٠٧	١٢	٠,٠٦٣
٥	٠,٠٢٤	١٣	٠,٠٢٢
٦	٠,٠٦١	١٤	٠,٠٠٥
٧	٠,١١٨	١٥	ــــ
			٠,٩٩٩

لحساب قوة الاختبار عندما يفترض أن ح = ٠,٧ نحسب احتمال وقوع قيم المتغير س في منطقة القبول وهي س ⩽ ١٢ تحت هذا الفرض أى على أساس أن للمتغير س توزيعا ذا حدين دليلاه ١٥ ، ٠,٧ .

احتمال الخطأ من النوع الثانى = β = ل(س ⩽ ١٢) حيث س : حد (١٥ ، ٠,٧)

= ١ − ل (س > ١٢)

من الجدول (٣) بذيل هذا الكتاب وبأخذ ن = ١٥ و ح = ٠,٧ نجد أن :

β = ١ − (٠,٠٩٢ + ٠,٠٣١ + ٠,٠٠٥) = ١ − ٠,١٢٨

= ٠,٨٧٢ = ٠,٨٧ تقريبا

وإذن قوة الاختبار = ى = ١ − ٠,٨٧ = ٠,١٣

ويلاحظ أن قوة الاختبار ضعيفة مما يدعونا إلى الشك فى قدرة التجربة على التمييز بين معدلى الشفاء فى العلاجين القياسى والجديد . وينبغى حينئذ العمل على زيادة هذه القدرة وذلك بزيادة حجم العينة .

حل آخر :

فى هذا المثال يمكننا استخدام تقريب التوزيع المعتدل لتوزيع ذى الحدين ـــ راجع البند (٤ − ٦) مع ملاحظة أن ن ح = ١٥ × ٠,٦ = ٩ > ٥ كما أن ن ك = ١٥ × ٠,٤ = ٦ > ٥ ، ويقترب توزيع ذى الحدين حد (١٥ ، ٠,٦) من توزيع معتدل متوسطه μ = ن ح = ٩ وانحرافه المعيارى σ = $\sqrt{\text{ن ح ك}}$ = $\sqrt{١٥ \times ٠,٦ \times ٠,٤}$ = ١,٨٩٧ وبالتالى يكون للمتغير

$$\text{ع}_\text{ن} = \frac{(\text{س} - ٠,٥) - ٩}{١,٨٩٧}$$

بالتقريب توزيع معتدل معيارى : مع (٠ ، ١) .

لإيجاد القيمة الحرجة أ لدينا :

$$ل\,(س > أ) = ل\,(ع > \frac{(أ - ٠{,}٥) - ٩}{١{,}٨٩٧}) = ٠{,}٠٥ \quad فرضا$$

من جدول المساحات أسفل المنحنى المعتدل المعيارى نجد أن

$$\frac{أ - ٩{,}٥}{١{,}٨٩٧} = ١{,}٦٤ \qquad ومنها\ أ = ١٢{,}٦١١$$

وإذن نرفض الفرض الصفرى أن ح = ٠,٦ عند مستوى الدلالة ٠,٠٥ إذا كان عدد المرضى الذين شفوا فى عينة من ١٥ مريضا يزيد عن ١٢. وهذه هى النتيجة التى توصلنا إليها بالحل الأول . كذلك :

β = ل (س ⩽ ١٢) حيث س : حد (١٥ ، ٠,٧)

وبالتقريب بتوزيع معتدل وسطه الحسابى ن ح = ١٥ × ٠,٧ = ١٠,٥

وانحرافه المعيارى σ = $\sqrt{ن\ ح\ ك}$ = $\sqrt{١٥ \times ٠{,}٧ \times ٠{,}٣}$ = ١,٧٧٤٨

نجد أن :

$$\beta = ل\,(س \leqslant ١٢) = ل\,(ع \leqslant \frac{(١٢ + ٠{,}٥) - ١٠{,}٥}{١{,}٧٧٤٨})$$

= ل (ع ⩽ ١,١٣) = ٠,٨٧٠٨ = ٠,٨٧ تقريبا

، ق = ١ − ٠,٨٧ = ٠,١٣

تمارين (٧)

(١) أخذت عينة عشوائية حجمها ن = ٥ من مجتمع معتدل تباينه ١٥,٢١ . بين كيف تختبر عند مستوى الدلالة ٠,٠٥ الفرض الصفرى أن متوسط المجتمع μ = ٤٥,٥ ضد الفرض $\mu \neq$ ٤٥,٥ . أوجد قوة الاختبار عندما نفترض أن μ = ٥٤ .

الفصل الثامن

تحليل التباين وتصميم التجارب

ANALYSIS OF VARIANCE
& DESIGN OF EXPERIMENTS

(٨ - ١) التحليل الإحصائي وتصميم التجارب :

يميل بعض الباحثين التجريبيين إلى تصميم تجاربهم وتنفيذها ، وبعد الانتهاء من الحصول على بيانات ينظرون في تحليل هذه البيانات إحصائياً . وهذا خطأ كبير لأن إغفال الجانب الإحصائي أثناء وضع التصميم غالباً ما يؤدى إلى اختيار تصميم خاطيء لا تستخلص منه أية نتائج يعتد بها . وعلى العكس من ذلك ، إذا أخذ الجانب الإحصائي بعين الاعتبار ، فإنه لا يعاون فقط على تحليل البيانات تحليلا علمياً سليماً بل يسهم بشكل أساسي في اختيار التصميم الأكثر كفاءة أى الذى يعطى أكبر قدر من المعلومات والنتائج بأدنى حد من الجهد التجريبي ، وهو بالإضافة إلى ذلك يقلل من مصادر أخطاء التجريب ويوضح طريقة تقدير هذه الأخطاء . فخطة التحليل الإحصائي هى جزء رئيسي من تصميم التجربة . وتقتضي هذه الخطة مراعاة عدة مبادىء لعل أهمها ما يلى :

(أولا) العشوائية : RANDOMNESS

إن التقنية الإحصائية للتجريب تقتضي تطبيق مبدأ العشوائية في كل ما يتعلق بالتجربة منعاً لأى تحيز من أى نوع ، وكوسيلة للتصدى لمجموعة العوامل الثانوية

التي نعجز عن حصرها أو حساب التأثير الطفيف الذى يحدثه كل منها وبالتالى نعجز عن التحكم فيها تجريبياً .

(أ) فالعينة التي تختار من المجتمع ينبغى أن تكون عشوائية ، فتكون مسحوبة بحسب خطة تضمن عدم وجود تحيز من أى نوع قد يؤثر فى عملية الاختيار .

(ب) وإذا قسمت هذه العينة إلى أقسام لتطبيق أنواع مختلفة من المعالجات على هذه الأقسام ينبغى أن يكون هذا التقسيم عشوائياً لكى يتوفر لكل وحدة من وحدات العينة نفس الفرصة لتلقي أى من هذه الأنواع .

(جـ) كما أن توزيع نوع ما لمعالجة ما على وحدات قسم ما ينبغى أن يكون عشوائياً خاصة من حيث الترتيب الزمني .

وبالنسبة لتقسيم العينة هناك طرق تكفل عشوائية هذا التقسيم ، ومن هذه الطرق ما يلى .

(١) رمى قطعة معدنية من العملة :

نفرض مثلا أننا نريد تقسيم ٢٠ حشرة إلى ٤ أقسام يحتوى كل منها على ٥ حشرات لكى تتلقي الحشرات التي تدخل في قسم ما واحداً من ٤ معالجات مختلفة ا ، ب ، حـ ، ى . نرقم الحشرات من ١ إلى ١٢٠ . نأخذ كل حشرة على حدة ونرمى قطعة منتظمة من العملة مرتين عشوائياً (أو نلقي قطعتين متميزتين من العملة مرة واحدة) . نحدد القسم الذى تدخل فيه الحشرة بحسب خطة كالآتية :

الرمية الأولى	الرمية الثانية	القسم (المعالجة)
صورة	صورة	ا
صورة	كتابة	ب
كتابة	صورة	حـ
كتابة	كتابة	ى

فإذا أخذنا الحشرة رقم (١) وظهرت كتابة في الرمية الأولى وصورة في الرمية الثانية فإن هذه الحشرة تدخل القسم حـ أى تتلقى المعالجة حـ وهكذا بالنسبة للحشرات جميعاً . وإذا حدث أن امتلأ أحد الأقسام (بخمس حشرات) وجاءت رمية لهذا القسم نلغى هذه الرمية ونعيد الرمى . إن هذه الخطة تكفل أن تكون النواتج الممكنة من رمى العملة مرتين وهى (صورة ، صورة) و(صورة ، كتابة) و(كتابة ، صورة) و(كتابة ، كتابة) متساوية الاحتمال إذ من الواضح أن احتمال كل منها يساوى $\frac{1}{4}$ بشرط أن تكون العملة منتظمة والرمى عشوائياً . وبهذا يكون لكل حشرة نفس الفرصة لتلقي أى من المعالجات الأربع . وهذا يعني توفر شرط عشوائية التقسيم . نلاحظ أنه بالنسبة للحشرة الأخيرة لا نكون بحاجة إلى رمى قطعة العملة .

(٢) رمي حجرة نرد :

في المثال السابق يمكن أن نستخدم خطة أخرى كالآتية :

نرمى حجرة نرد منتظمة عشوائيا . إذا ظهرت نقطة واحدة ندخل الحشرة في القسم أ وإذا ظهرت نقطتان ندخلها في القسم ب وإذا ظهرت ٣ نقط ندخلها في القسم حـ وإذا ظهرت ٤ نقط ندخلها في القسم ء . أما إذا ظهرت ٥ أو ٦ نقط فلا تحسب ويعاد الرمى . نلاحظ هنا أيضاً أن النواتج الستة متساوية الاحتمال فاحتمال كل منها يساوى $\frac{1}{6}$.

(٣) استخدام ورق اللعب :

حين يكون عدد الأقسام المطلوبة كبيراً يحسن استخدام ورق اللعب . نفرض أننا نريد تقسيم ٦٠ حشرة إلى ١٢ قسماً يحتوى كل منها على ٥ حشرات لكى تتلقى الحشرات التى تدخل فى قسم ما واحداً من ١٢ معالجة مختلفة . نستخدم مجموعة من ورق اللعب بعد استبعاد الملوك الأربعة فيكون لدينا ٤٨ ورقة . نحدد خطة كالآتية :

الواحد للقسم الأول والاثنين للقسم الثاني ، ... ، ... والعشرة للقسم العاشر والولد للقسم الحادى عشر والبنت للقسم الثاني عشر . نأخذ كل حشرة على حدة ونخلط الورق جيداً ثم نقطعه عشوائياً فيكون العدد المقطوع هو الذى يحدد القسم الذى تدخل فيه الحشرة . نلاحظ أن احتمال ظهور أى من الحالات الاثني عشر يساوى $\frac{٤}{٤٨} = \frac{١}{١٢}$.

إن الطرق سابقة الذكر هى مجرد أمثلة على طرق التقسيم العشوائي ويمكن للباحث على ضوء هذه الأمثلة أن يبتدع طرقاً أخرى كثيرة ، وذلك إضافة إلى إمكانية استخدام جداول الأعداد العشوائية المشار إليها بالبند (١ – ٢) .

(ثانيا) الاستقلال : INDEPENDENCE

يقتضي التحليل الإحصائي ، خاصة في تحليل التباين ، افتراض استقلال أخطاء التجريب عن بعضها واستقلال المعالجات عن أخطاء التجريب ، فإذا كانت الأخطاء الكبيرة مثلا مرتبطة بمعالجة معينة فإن استخدام تقدير شامل لخطأ التجريب ، وهو الإجراء المتبع عادة ، لا يكون إجراء سليماً يعتمد عليه في اختبارات الدلالة . على أن تطبيق مبدأ العشوائية سابق الذكر يضمن إلى حد كبير تحقيق هذا الافتراض . كما يسهم في تحقيقه استخدام مجموعات من المشاهدات (مأخوذة من سلسلة من التجارب) بدلا من استخدام مجموعة واحدة من المشاهدات .

(ثالثا) النموذج الإحصائي : STATISTICAL MODEL

كما يقتضي التحليل الإحصائي وضع نموذج يرشدنا إلى الأسس الإحصائية التي تحدد أسلوب هذا التحليل ، وتعكس حدوده تأثيرات العوامل أو المتغيرات التي تدخل في التجريب . وترتبط بكل نموذج افتراضات خاصة تتعلق بتوزيعات هذه المتغيرات أو باستقلالها أو بالصورة الرياضية التي يأخذها النموذج لوصف وحدات التجريب ... وكلما كان النموذج ناجحاً في تصوير التجربة الفعلية كلما كانت النتائج التي نحصل عليها من تحليل البيانات أكثر صدقاً .

(رابعاً) مسائل أخرى :

ينبغى أن يجيب تصميم التجربة على تساؤلات عدة منها :

(أ) ما هو الحجم المناسب للعينة ؟

(ب) متي يتعين تكرير التجربة برمتها ؟

(جـ) متي نحتاج إلى إدخال مجموعة مراقبة ؟ control group

(د) ما الطريقة العملية لتطبيق مبدأ العشوائية ؟

(هـ) ما مدى الدقة والضبط اللازمين في عملية القياس ؟

(٨ - ٢) تحليل التباين :

كثيراً ما نلاحظ وجود اختلاف في قيم متغير ما لا نعرف سببه أو مصدره ولا نستطيع التحكم فيه . ومن أمثلة ذلك الإختلاف المشاهد في الزيادة الشهرية في أوزان مجموعة من الماشية حتي لو وضعت في ظروف واحدة وتحت نظام غذائي مشترك ، كذلك الاختلاف المشاهد في نمو وحدات نبات مزروع في حقل تحت نفس الظروف ... إن مثل هذا الاختلاف نصفه بأنه **اختلاف عشوائي** .

على أننا في كثير من التجارب ندخل سبباً إضافياً للاختلاف في قيم المتغير نعلم مصدره ، فمثلا قد نقسم مجموعة الماشية إلى عدة أقسام يتلقي كل منها نظاماً مختلفاً للتغذية ، أو قد يقسم الحقل إلى عدة أحواض يتلقي كل منها نوعاً مختلفاً من المخصبات أو طرقاً مختلفة للرى ونقول حينئذ أننا أدخلنا عاملا factor معيناً في التجربة . والعامل هو متغير نوعى يتألف من عدد من المعالجات treatments أو التقسيمات المرتبطة تسمى مستويات العامل levels فالعامل « نظام التغذية » قد يتكون من ٤ مستويات والعامل « نوع المخصب » قد يتكون من ٣ مستويات وهكذا ..

ومن الواضح أن الهدف من إدخال العامل معرفة ما إذا كانت المستويات المختلفة (لنظام التغذية مثلا) تحدث تأثيرات مختلفة في قيم المتغير (الزيادة في الوزن)

وهذا هو الغرض الذى تستخدم من أجله عملية تحليل التباين . وتؤسس هذه العملية على تصميم تجربة تمكننا من أن نفصل الاختلاف الذى سببه العامل عن الاختلاف العشوائي ، فإذا ظهر لنا أن ذلك الاختلاف جوهرى حكمنا بأن عامل التقسيم هو عامل مؤثر في قيم المتغير وتصدينا بعد ذلك للمقارنة بين مستويات هذا العامل .

فتحليل التباين هو عملية نستطيع بواسطتها أن نحلل الاختلاف الكلى المشاهد في مجموعة من البيانات إلى مركبتين أو أكثر يرجع كل منها إلى عامل أو مصدر مستقل ، وإذا كانت هذه البيانات من عينة عشوائية مأخوذة من مجتمع معتدل ذى تباين معين فإن كلا من هذه المركبات يعطى تقديراً مستقلا لهذا التباين ، والتقديرات الناتجة يمكن اختبار تجانسها بواسطة اختبار ف .

وفي التجارب التي نبحث فيها تأثير عامل واحد – كما في الأمثلة سابقة الذكر – يحلل الاختلاف الكلى إلى مركبتين مستقلتين إحداهما تناظر هذا العامل والأخرى تناظر الاختلاف العشوائي . وفي التجارب التي نبحث فيها تأثير عاملين مستقلين ، مثلا نوع الغذاء كأحد العاملين وكمية الغذاء كعامل ثان ، نحلل الاختلاف الكلى إلى ثلاث مركبات مستقلة اثنتان منهما تناظران العاملين والثالثة تناظر الاختلاف العشوائي ، وهكذا في حالة وجود أكثر من عاملين .

وفي بحثنا عن كيفية تحليل التباين نحتاج إلى المصطلحات والتعاريف المبينة في البند التالى .

(٨ – ٣) مصطلحات وتعاريف :

(أ) مجموع المربعات (م م) : SUM OF SQUARES (SS)

إذا كان لدينا مجموعة من القيم $س_1$ ، $س_2$ ، . . . ، $س_ن$ فإن مجموع مربعات

انحرافات هذه القيم عن وسطها الحسابي $\bar{س}$ يسمى اختصارا بمجموع المربعات ونرمز له بالرمز م م ، أى أن :

$$م\,م = \text{مجـ}_{ر=1}^{ن} (س_ر - \bar{س})^2$$

$$= \text{مجـ}_ر\, س_ر^2 - \frac{ج^2}{ن}$$ حيث $ج = \text{مجـ}_ر\, س_ر$ = مجموع القيم .

ويتخذ م م كمقياس للاختلاف variation في هذه القيم . ويلاحظ أن قيمة م م لا تتغير إذا جمعنا أو طرحنا عددا ثابتا من جميع القيم .

(ب) درجات الحرية (ر) أو (د.ح)

DEGREES OF FREEDOM

يستخدم مصطلح « درجات الحرية » فى الإحصاء التطبيقى للتعبير عن عدد المقادير المستقلة خطيا فى مجموع المربعات . فمثلا القيم $س_1$ ، $س_2$ ، . . . ، $س_ن$ هى ن من المقادير المستقلة ولكن نظرا لأن $\text{مجـ}\,(س_ر - \bar{س}) = 0$. فإن المقادير $(س_ر - \bar{س})$ لا تكون مستقلة خطيا لأنه يمكن اشتقاق واحد منها من الآخرين ولذلك يكون لمجموع مربعاتها $\text{مجـ}\,(س_ر - \bar{س})^2$ درجات حرية عددها ن − ١ .

كقاعدة عامة ، يحسب عدد درجات الحرية لإحصاءة ما كالآتي :

= عدد المشاهدات المستقلة المسببة للاختلاف − عدد البارامترات المستقلة التي قدرت من العينة عند حساب هذا الاختلاف .

وفى تحليل التباين نستخدم التعريف الإجرائي الآتي لدرجات الحرية لأى مصدر من مصادر الاختلاف .

= عدد الانحرافات المربعة − عدد النقط (المحاور) المستقلة التي أخذت حولها هذه الانحرافات . (يلاحظ أن عدد النقط أو المحاور هذه هى عدد القيود الخطية التي فرضت على تقدير الاختلاف) .

(حـ) متوسط المربعات ($ع^٢$) أو (ط م) MEAN SQUARE (MS)

هو خارج قسمة مجموع المربعات على عدد درجات الحرية أى $ع^٢$= م م / د ويسمى هذا بالتباين ، غير أن التعبير متوسط المربعات هوتعبير أكثر عمومية .

(٨ - ٤) التجارب ذوات العامل الواحد :

SINGLE FACTOR EXPERIMENTS

في هذه التجارب يكون اهتمامنا منصبا على دراسة عامل واحد فقط ، وليكن نظام التغذية ، من حيث تأثيره على متغير ما وليكن الزيادة في وزن نوع من البقر في مدة ما . ولا يغرب عن بالنا هنا إمكانية وجود مصادر أو عوامل أخرى ذات تأثير على هذا المتغير مثل عمر البقر وجنسه ووزنه الأصلى .. ولذلك ينبغى أن نعمل على تحييد تأثير هذه العوامل ومنع تداخل هذا التأثير مع التأثير الذى يحدثه العامل الذى ندرسه .

ولتحقيق هذا الغرض يلجأ بعض الباحثين إلى تصميم تجربة يتحكم فيها تحكما كاملا في هذه العوامل فيقوم بتثبيتها عند مستويات محددة فيختار مجموعة من البقر في نفس العمر ومن نفس الجنس ونفس الوزن .. ويسمح فقط بتغيير عامل التغذية وذلك بتقسيم مجموعة البقر إلى عدة أقسام ومعالجة كل قسم بواحد من مستويات نظام التغذية ، وبهذا يخلى مسئولية أى من تلك العوامل مما قد يظهر من فروق جوهرية بين هذه المستويات . غير أن النتائج التى تسفر عنها هذه الطريقة تكون مشروطة بتوفر الظروف الخاصة التي هيئت لها التجربة من حيث العمر والجنس والوزن .. وقد لا تكون هذه النتائج صحيحة إذا ما تغير أى من هذه الظروف ، وبالتالى لا تعطى التجربة قدرا كافيا من المعلومات التي ينشدها الباحث . ومن ناحية أخرى من الصعب عمليا بل ومن المستحيل أحيانا التحكم في التجربة وضبط تلك العوامل الخارجية عند مستوى مشترك بالدقة الكافية .

ولذلك يفضل الباحثون تصميم تجربة على النقيض من ذلك ، فبدلا من أن نتحكم في العوامل الخارجية بوضعها عند مستويات خاصة ، نقوم **بتعشية هذه** العوامل بحيث يكون لكل مستوى من مستويات العامل الذى ندرسه نفس الفرصة للتعرض لها ، وبهذا يحق لنا ضم تأثير هذه العوامل تحت كلمة عامة هى **الاختلاف العشوائي** أو خطأ التجريب أو الصدفة . فإذا كان لدينا ٢٨ بقرة و٤ مستويات من نظام التغذية نرقم البقر من ١ إلى ٢٨ بصرف النظر عن العمر والجنس والوزن لتقسيم البقر عشوائيا إلى ٤ أقسام يحتوى كل منها على ٧ بقرات لتتلقي واحدا من مستويات نظام التغذية . إن مثل هذا التصميم يسمى **بالتصميم كامل التعشية** completely randomized design وهو ذلك التصميم الذى يبني لا على أساس التحكم في المصادر الخارجية وإنما على أساس تعشية هذه المصادر بحيث يمكن ضم الاختلافات الناشئة عنها تحت اسم **الاختلاف العشوائي** .

على أن طريقة التعشية في الحماية من تأثير العوامل الخارجية هي عملية مبنية على أساس احتمالى ، فقد تسفر هذه الطريقة عن أن يشتمل أحد الأقسام على ٧ بقرات كلها من الإناث أو كلها من صغار السن – وإن كان هذا أمرا بعيد الاحتمال – وهذا أحد الأسباب التي تجعل بعض الباحثين يميل إلى استخدام تصميم وسط بين النقيضين المذكورين وذلك بالتحكم في بعض العوامل وتعشية البعض الآخر . وسنعود إلى هذا الموضوع فى البند (٨ — ٧) تحت عنوان المقارنات التزاوجية.

اعتبر عينة عشوائية حجمها ں مأخوذة من متغير معتدل س وسطه الحسابى μ وتباينه σ^2 . افرض أن هذه العينة قسمت عشوائياً إلى ك من الأقسام تلقت كل منها واحداً من مستويات عامل ما (نظام الغذاء – نوع المخصب – طريقة الرى – ...) وجاءت البيانات كما يلي ، حيث س~ر~ه~ ترمز إلى قيم المتغير الذى ندرسه (مقدار المحصول مثلا) ، وحيث ر ، ه ترمزان على الترتيب إلى رقم الصف ورقم العمود الذى تقع فيه القيمة س~ره~.

	الأقسام (المعالجات)							
	(١)	(٢)	(٣)	...	(ق)	...	(ك)	
	$س_{١١}$	$س_{٢١}$	$س_{٣١}$	...	...	...	$س_{١ك}$	
	$س_{١٢}$	$س_{٢٢}$	$س_{٣٢}$	...	...	...	$س_{٢ك}$	
	$س_{١٣}$	$س_{٢٣}$	$س_{٣٣}$	...	...	...	$س_{٣ك}$	
	...	...	...	...	...	...	...	
	...	...	...	...	...	...	...	
	$س_{ر١}$	$س_{ر٢}$	$س_{ر٣}$	...	$س_{رق}$	...	$س_{رك}$	
	...	...	...	...	...	...	...	
	$س_{ن_١ ١}$	$س_{ن_٢ ٢}$	$س_{ن_٣ ٣}$	...	...	...	$س_{ن_ك ك}$	
$ن_ق$	$ن_١$	$ن_٢$	$ن_٣$	...	$ن_ق$	...	$ن_ك$	$ن = مجـ ن_ق$
$م_ق$	$م_١$	$م_٢$	$م_٣$	...	$م_ق$	...	$م_ك$	$م = مجـ م_ق$
$\bar{س}_ق$	$\bar{س}_١$	$\bar{س}_٢$	$\bar{س}_٣$	...	$\bar{س}_ق$	...	$\bar{س}_ك$	$\bar{س} = مجـ م / ن$

$ن_ق$ ترمز إلى عدد قيم المتغير فى القسم ق (ق = ١ ، ٢ ، ... ، ك)

، ن ترمز إلى العدد الكلى لقيم المتغير

، $م_ق$ ترمز إلى مجموع قيم المتغير فى القسم ق

، $\bar{س}_ق$ ترمز إلى متوسط قيم المتغير فى القسم ق ، $\bar{س}$ ترمز إلى المتوسط العام.

المطلوب بحث ما إذا كان المجتمع متجانساً بالنسبة لهذا التقسيم أى ما إذا كانت مستويات هذا العامل تحدث تأثيرات متساوية في قيم المتغير س .

(٨ – ٤ – ١) النموذج الإحصائي (النموذج I) :

كما سبق القول يعتمد التحليل الإحصائي على اختيار نموذج يعبر عن تركيب أي عنصر مشاهد في التجربة ويبرر ما يجرى من عمليات مصحوباً بافتراضات يقتضيها البناء الرياضي الذى تقوم عليه عملية التحليل . وسنفترض هنا ما يلي :

(١) المجتمع العام الذى أخذت منه وحدات التجريب هو مجتمع **معتدل** وسطه الحسابي μ وتباينه $\sigma^{٢}$.

(٢) مجموعات الوحدات في الأقسام ١ ، ٢ ، ... ، ك التي تلقت مستويات مختلفة من عامل التجريب تشكل عينات عشوائية **مستقلة** مأخوذة من ك من المجتمعات **المعتدلة** أوساطها الحسابية $\mu_{١}$ ، $\mu_{٢}$ ، ... ، $\mu_{ك}$ ولها تباين مشترك $\sigma^{٢}$ ($\sigma_{١}^{٢} = \sigma_{٢}^{٢} = ... = \sigma_{ك}^{٢} = \sigma^{٢}$) .

(٣) أى وحدة مشاهدة $س_{رو}$ تخضع للنموذج الخطى الآتي :

$$س_{رو} = \mu_{و} + خ_{رو} \qquad ر = ١ ، ٢ ، ... ، ن_{و} \qquad (١)$$

حيث $\mu_{و}$ هو الوسط الحسابي لمجتمع القسم و (و = ١ ، ٢ ، ... ، ك) . ، $خ_{رو} = س_{رو} - \mu_{و}$ هو خطأ التجريب بالنسبة للوحدة $س_{رو}$ التي في الصف ر والعمود (القسم) و أى أن قيم $خ_{رو}$ تعبر عن الفروق العشوائية بين الوحدات داخل القسم و ، وسنفرض أنه في مجتمع القيم $خ_{رو}$ تكون هذه القيم مستقلة ويكون هذا المجتمع معتدلا وسطه الحسابي صفر وتباينه $\sigma^{٢}$.

ومن المعتاد أن يكتب النموذج (١) كالآتي :

$$س_{رو} = \mu + \alpha_{و} + خ_{رو} \qquad (٢)$$

حيث $\alpha_{و} = \mu_{و} - \mu$ = انحراف متوسط مجتمع القسم و عن متوسط المجتمع العام ، وهذا الرمز يصلح للتعبير عن متوسط أثر المستوى و لعامل التقسيم .

وسنعتبر أن هذا الأثر ثابت لكل وحدة بالقسم ق وأنه يختلف من قسم إلى آخر ، بمعنى أن كل عنصر من عناصر القسم الأول يتأثر (بالزيادة أو النقصان) بمقدار ثابت α_1 وكل عنصر من عناصر القسم الثانى يتأثر بمقدار ثابت α_2 وهكذا ... ولذلك يسمى هذا النموذج بالنموذج ثابت التأثيرات fixed effects model تمييزا له عن النموذج عشوائى التأثيرات random effects model حيث لا تتأثر عناصر الأقسام بمقادير ثابتة بل بمقادير عشوائية . وسوف نتناول هذا النموذج فى البند (٨ − ١٤) .

إن هذا النموذج هو الأساس الذى يبنى عليه التحليل ، فهو أولا يفترض أن أى قيمة مشاهدة $\text{س}_{\text{رو}}$ يمكن تجزئتها إلى مركبات تعزى إلى مصادر منطقية متميزة نتبينها كالآتى :

من (٢) : $$\text{س}_{\text{رو}} = \mu + (\mu_{\text{و}} - \mu) + (\text{س}_{\text{رو}} - \mu_{\text{و}}) \qquad (٣)$$

ومن هذه الصيغة نرى أن $\text{س}_{\text{رو}}$ وهى القيمة المشاهدة فى الصف رَ من القسم و تساوى المتوسط العام للمجتمع + تأثير يرجع إلى المعالجة التى تلقتها وحدات القسم و + تأثير عشوائى داخل القسم و .

كما أن هذا النموذج يحدد العلاقة بين الاختلافات الناشئة عن مختلف المصادر أو العوامل المؤثرة فى عملية التجريب :

من (٣) : $$\text{س}_{\text{رو}} - \mu = (\mu_{\text{و}} - \mu) + (\text{س}_{\text{رو}} - \mu_{\text{و}})$$

أو $$(\text{س}_{\text{رو}} - \bar{\bar{\text{س}}}) = (\bar{\text{س}}_{\text{و}} - \bar{\bar{\text{س}}}) + (\text{س}_{\text{رو}} - \bar{\text{س}}_{\text{و}}) \qquad (٤)$$

وذلك بوضع المتوسطات $\bar{\bar{\text{س}}}$ ، $\bar{\text{س}}_{\text{و}}$ الناتجة فى العينة بدلا من المتوسطات النظرية المجهولة μ ، $\mu_{\text{و}}$. بتربيع كل من الطرفين والجمع على جميع قيم ر ، و تنتج المتطابقة الآتية :

$$\underset{\text{و}}{\text{مج}}\ \underset{\text{ر}}{\text{مج}}(\text{س}_{\text{رو}} - \bar{\bar{\text{س}}})^2 = \underset{\text{و}}{\text{مج}}\ \text{ن}_{\text{و}}(\bar{\text{س}}_{\text{و}} - \bar{\bar{\text{س}}})^2 + \underset{\text{و}}{\text{مج}}\ \underset{\text{ر}}{\text{مج}}(\text{س}_{\text{رو}} - \bar{\text{س}}_{\text{و}})^2 \qquad (٥)$$

وهذه العلاقة صحيحة دائما سواء كانت المجتمعات معتدلة أو غير معتدلة ، وهى العلاقة الأساسية فى تحليل التباين ، وتشير إلى أن الاختلاف الكلى فى بيانات التجربة وهو مجـ مجـ $(س_{رى} - \bar{س})^2$ يتحلل إلى المركبتين الآتيتين :

(ا) المركبة مجـ $ن_ى (\bar{س}_ى - \bar{س})^2$

وهى تعبر عن الاختلاف بين متوسطات الأقسام (مرجحة بأعداد عناصر هذه الأقسام) ويرجع هذا الاختلاف بالطبع إلى عامل التقسيم ، أى إلى اختلاف تأثير مستويات عامل التجريب على قيم المتغير س . ونرمز لهذا الاختلاف بالرمز م م (بين الأقسام) وعدد درجات حريته $\nu_2 = ك - ١$ لأن هناك ك من الانحرافات المربعة وأخذت جميعها حول محور واحد هو $\bar{س}$.

(ب) المركبة مجـ مجـ $(س_{رى} - \bar{س}_ى)^2$

وهى تعبر عن مجموع الاختلافات العشوائية داخل الأقسام ، مع ملاحظة أن لكل قسم ى اختلاف عشوائى قدره مجـ$_ر$ $(س_{رى} - \bar{س}_ى)^2$ بدرجات حرية $ن_ى - ١$ وإذن مجموع الاختلافات العشوائية داخل الأقسام كلها هو مجـ$_ى$ مجـ$_ر$ $(س_{رى} - \bar{س}_ى)^2$ بدرجات حرية عددها ν_3 = مجـ$_ى$ $(ن_ى - ١) = ن - ك$. ونرمز لهذا الاختلاف بالرمز م م (داخل الأقسام) .

وإذا رمزنا للاختلاف الكلى بالرمز م م (الكلى) بدرجات حرية $\nu_1 = ن - ١$ فإن المتطابقة (٥) تكتب كالآتى :

الاختلاف الكلى = الاختلاف بين الأقسام + الاختلاف داخل الأقسام .
أى م م (الكلى) = م م (بين الأقسام) + م م (داخل الأقسام) .

حيث $\nu_1 = \nu_2 + \nu_3$ لأن $ن - ١ = (ك - ١) + (ن - ك)$

من هذا نحصل على متوسط المربعات لكل من المركبتين كالآتي :

$$ع^2_ب = م م (بين الأقسام) / (ك - ١)$$

، $$ع^2_خ = م م (داخل الأقسام) / (ن - ك)$$

تحت الفروض سابقة الذكر نستطيع أن نثبت رياضياً أن $ع^2_ب$ ، $ع^2_خ$ هما تقديران مستقلان لتباين المجتمع σ^2 . غير أن التقدير الثاني ($ع^2_خ$) هو تقدير غير متحيز لتباين المجتمع بمعني أن متوسط مثل هذه التقديرات يكون على المدى البعيد مساوياً للتباين σ^2 ، أما التقدير الأول ($ع^2_ب$) فهو تقدير متحيز للتباين σ^2 إذ أن متوسط مثل هذه التقديرات على المدى البعيد $\sigma^2 + \frac{1}{ك - 1} \sum ن_ر (\mu_ر - \mu)^2$ أى يزيد عنه ولا يكون مساوياً له إلا إذا كانت المتوسطات μ_1 ، μ_2 ، ... ، $\mu_ك$ متساوية جميعا .

وعلى ذلك فإن أى فرق جوهرى بين التقديرين $ع^2_ب$ ، $ع^2_خ$ لا ينتج إلا من وجود فرق جوهرى بين هذه المتوسطات .

(٨ - ٤ - ٢) اختبار تجانس المجتمع بالنسبة لعامل التقسيم :

إن الفرض الصفرى هو أن المجتمع متجانس بالنسبة لعامل التقسيم وهذا يتضمن أن يكون للمجتمعات المعتدلة للأقسام نفس الخواص الإحصائية ، وبصفة خاصة :

$$\mu_1 = \mu_2 = \dots = \mu_ك ، \sigma^2_1 = \sigma^2_2 = \dots = \sigma^2_ك = \sigma^2$$

وإذا كان هذا الفرض صحيحاً ، ونظراً لأن $ع^2_ب$ ، $ع^2_خ$ هما تقديران مستقلان للتباين σ^2 لمجتمع معتدل فإن نسبة التباين

$$ف = \frac{ع^2_ب}{ع^2_خ} \qquad (٦)$$

يكون توزيعها مطابقاً لتوزيع المتغير ف بدرجتي حرية (ك - ١ ، ن - ك) - راجع البند (٦ - ٨) . وعلى ذلك يمكن استخدام اختبار ف للحكم على هذا التجانس وبالتالى للحكم على تساوى تلك المتوسطات .

أما الفرض الآخر $ف_1$ فهو أن الاختلاف الناشيء عن عامل التقسيم (بين المتوسطات) أكبر مما نتوقعه من اختلاف عشوائي في عينة من مجتمع متجانس بالنسبة لعامل التقسيم ، ولذلك فإن هذا الاختبار **يكون دائماً ذا جانب واحد** .

(٨ — ٤ — ٣) طريقة مختصرة لحساب الاختلاف :

لتسهيل حساب القيم العددية لمجاميع المربعات الثلاثة المبينة بالمتطابقة (٥) نستخدم الصيغ الآتية التي يمكن برهنتها رياضياً .

(١) م م (الكلي) $= \text{مجـ}_و \text{مجـ}_ر س^2_{رو} - \frac{م^2}{ن}$ حيث $\nu_1 = ن - 1$

(٢) م م (بين الأقسام) $= \text{مجـ}_و \frac{م^2_و}{ن_و} - \frac{م^2}{ن}$ حيث $\nu_2 = ك - 1$

(٣) م م (داخل الأقسام) = (١) — (٢) حيث $\nu_3 = ن - ك$

وتوضع هذه القيم عادة في جدول يسمى بجدول التباين يأخذ الصورة الآتية :

الجدول (٨ — ٢)

جدول التباين للتجارب ذوات العامل الواحد

مصدر التباين	م م	د ح	تقدير التباين	$ف_ب$
بين الأقسام	(٢)	ك — ١	$ع^2_ب$	$ع^2_ب / ع^2_خ$
داخل الأقسام	(١) — (٢)	ن — ك	$ع^2_خ$	
المجموع	(١)	ن — ١		

ملاحظة (١) :

يفضل أن تكون حجوم العينات في الأقسام المختلفة متساوية أى $ن_١ = ن_٢ = ... = ن_ه = ا$ مثلا لأنه في هذه الحالة لا يكون تحليل التباين حساساً للانحرافات الصغيرة عن فرض تساوى التباينات في مجتمعات الأقسام . هذا بالإضافة إلى تسهيل حساب مجاميع المربعات بين الأقسام حيث تكتب كما يلى :

$$م^٢ \text{ (بين الأقسام) } = \frac{مج\, م_ه^٢}{ا} - \frac{م^٢}{ن}$$

ملاحظة (٢) :

لا تتأثر نتائج تحليل التباين بأى حال إذا جمعنا أو طرحنا عدداً ثابتاً من جميع قيم وحدات التجريب .

مثال (٨ — ١) :

البيانات التى بالجدول (٨ — ٣) نتجت عن تجربة في فسيولوجيا النبات ، وهى تعطى الطول (بوحدات شفرية) لمقاطع من نبات البسلة تركت لتنمو في مزرعة نسيجية في وجود هرمون الأوكسين ، وكان الهدف من التجربة اختبار تأثير إضافة أربعة أنواع من السكريات على النمو مقاساً بواسطة الطول .

(على فرض أن مبادىء العشوائية والاعتدالية قد روعيت في إجراء التجربة .)

الحل :

الفرض الصفرى ف هو أن المجتمع (المعتدل) الذى أخذت منه العينة متجانس بالنسبة لعامل التقسيم (نوع السكر) أى أن $\mu_١ = \mu_٢ = \mu_٣ = \mu_٤ = \mu_٥$. والفرض الآخر $ف_١$ هو : على الأقل اثنان من المتوسطات غير متساويين .

جدول (٨ - ٣)

	المعالجات (١) +٢٪ جلوكوز	(٢) +٢٪ فركتوز	(٣) +١٪ جلوكوز +١٪ فركتوز	(٤) +٢٪ سكروز	(٥) مراقبة	
	٥٧	٥٨	٥٨	٦٢	٧٥	
	٥٨	٦١	٥٩	٦٦	٦٧	
	٦٠	٥٦	٥٨	٦٥	٧٠	
	٥٩	٥٨	٦١	٦٣	٧٥	
	٦٢	٥٧	٥٧	٦٤	٦٥	
	٦٠	٥٦	٥٦	٦٢	٧١	
	٦٠	٦١	٥٨	٦٥	٦٧	
	٥٧	٦٠	٥٧	٦٥	٦٧	
	٥٩	٥٧	٥٧	٦٢	٧٩	
	٦١	٥٨	٥٩	٦٧	٦٨	
نى	١٠	١٠	١٠	١٠	١٠	ن = ٥٠
مجى	٥٩٣	٥٨٢	٥٨٠	٦٤١	٧٠١	مج = ٣٠٩٧
س̄ى	٥٩,٣	٥٨,٢	٥٨	٦٤,١	٧٠,١	س̄ = ٦١,٩٤

م م (الكلي) = ٥٧² + ٥٨² + ... + ٥٨² + ٦١² + ... + ٥٨² + ٥٩² + ... + ٦٢² + ٦٦² + ... + ٧٥² + ٦٧² + ... + ٦٨²

$$- \frac{(٣٠٩٧)^2}{٥٠}$$

= ١٩٣١٥١ − ١٩١٨٢٨,١٨ = ١٣٢٢,٨٢

حيث ν_1 = ن − ١ = ٤٩

$م^2$ (بين الأقسام) = $\frac{(٥٩٣)^2}{١٠} + \frac{(٥٨٢)^2}{١٠} + ... + \frac{(٧٠١)^2}{١٠} - \frac{(٣٠٩٧)^2}{٥٠}$

= ١٩٢٩٠٥,٥٠ − ١٩١٨٢٨,١٨ = ١٠٧٧,٣٢

حيث ν_2 = ك − ١ = ٤

$م^2$ (داخل الأقسام) = ١٣٢٢,٨٢ − ١٠٧٧,٣٢ = ٢٤٥,٥٠

حيث ν_3 = ن − ك = ٤٥

نضم هذه النتائج في جدول التباين الآتي :

جدول (٨ − ٤)

مصدر الاختلاف	مجموع المربعات	درجات الحرية	تقدير التباين	ف ى
بين الأقسام (المعالجات)	١٠٧٧,٣٢	٤	٢٦٩,٣٣	٤٩,٣٣
داخل الأقسام (خطأ التجريب)	٢٤٥,٥٠	٤٥	٥,٤٦	
المجموع	١٣٢٢,٨٢	٤٩		

من جدول ف ، وعند درجتي الحرية ٤ ، ٤٥ نجد أن :

$ف_{٠,٠٥}$ = ٢,٥٨ ، $ف_{٠,٠١}$ = ٣,٧٧ ، $ف_{٠,٠٠١}$ = ٥,٥٧

الاستنتاج :

نظرا لأن ف ى = ٤٩,٣٣ أكبر بكثير من أى من هذه القيم فإننا نرفض الفرض الصفرى عند مستوى عالى من الدلالة ونحكم بأن الأنواع المختلفة من السكريات ليستمتساوية في تأثيرها على نمو مقاطع نبات البسلة .

ملاحظة (٣) :

في نسبة التباين ف نضع $ع^{2}_{ب}$ دائماً في البسط و$ع^{2}_{غ}$ في المقام وإذا حدث أن كانت $ع^{2}_{ب}$ أصغر من $ع^{2}_{غ}$ أى كانت ف $<$ ١ نقبل الفرض الصفرى فوراً دون حاجة إلى إيجاد أى قيمة حرجة من الجدول لأن جميع هذه القيم أكبر من الواحد حين تزيد كل من درجتي الحرية عن الواحد .

ملاحظة (٤) :

حين يكون عدد الأقسام ك = ٢ تكون نتيجة تحليل التباين مطابقة للنتيجة التي نحصل عليها باستخدام اختبار ت وتكون ف (١ ، ن − ١) = ت² (ن − ١) . ولذلك يمكن أن نعتبر أن اختبار ت للفرق بين متوسطى مجتمعين معتدلين هو حالة خاصة من اختبار ف .

(٨ − ٥) المقارنة بين المتوسطات :

في البند السابق أجرينا تحليلا للتباين المشاهد في البيانات التي أسفرت عنها التجربة ، غير أن هذا التحليل ليس إلا الخطوة الأولى لدراسة نتائج التجربة وينبغى أن يستكمل بإجراء مقارنات بين بعض أزواج هذه المتوسطات أو بين مجموعات منها . ودراسة هذه المقارنات قد يكون أكثر أهمية من التحليل العام . وهناك نوعان من المقارنات هما :

(أ) المقارنات القَبْلية — A priori (or planned) comparisons

(ب) المقارنات البَعْدية — A postiori (or unplanned) comparisons

ولعل سبب التمييز بين هذين النوعين هو اختلاف اختبارات الدلالة فيهما كما سيتبين بعد .

(٨ – ٥ – ١) الاختبارات القَبْلية :

هى تلك الاختبارات التي كان مخططاً لها أثناء تصميم التجربة (وقبل إجرائها) . ففي المثال (٨ – ١) كان مخططاً لاختبار تأثير إضافة السكريات ضد مجموعة المراقبة ، كما كان مخططاً لاختبار ما إذا كانت السكريات النقية ككل (جلوكوز – فركتوز – سكروز) تختلف في تأثيرها عن السكريات المختلطة (١٪ جلوكوز + ١٪ فركتوز) .

إن مثل هذه الاختبارات تجرى بصرف النظر عن النتيجة العامة لتحليل التباين أى سواء رفضنا أو قبلنا الفرض الصفرى عن تساوى المتوسطات .

والقاعدة التي نتبعها للمقارنة هى نفس القاعدة العامة وليس علينا إلا مراعاة أن نتناول فقط البيانات التي بالأقسام التي نرغب في مقارنتها ، وأن ننسب تقدير التباين الناتج منها إلى تقدير التباين (داخل الأقسام) السابق إيجاده في التحليل العام وهو $ع^{٢}_{ع}$ لأن هذا التقدير مبني على جميع ما لدينا من بيانات (وليس على جزء منها) فهو أقدر على تقدير تباين المجتمع $\sigma^{٢}$.

والأمثلة الآتية هى استكمال لدراسة التجربة التي بالمثال (٨ – ١) .

مثال (٨ – ٢) مقارنة مجموعة المراقبة ضد المجموعات الأخرى :

ابحث ما إذا كانت إضافة السكريات تؤثر في نمو مقاطع البسلة .

الحل :

الفرض الصفرى هو أن متوسط مجموعات السكريات الأربعة مجتمعة يساوى متوسط مجموعة المراقبة . نعتبر أن لدينا قسمين يتألف الأول من الأقسام ١ ، ٢ ، ٣ ، ٤ وبها ٤٠ عنصراً ، مجموع قيمها ٢٣٩٦ ، ومتوسطها ٥٩,٩ ، ويتألف

الثاني من قسم المراقبة وبه ١٠ عناصر مجموع قيمها ٧٠١ ومتوسطها ٧٠,١ ويكون المجموع الكلى لقيم العناصر التي اخترناها لهذه الدراسة ٣٠٩٧ .

$$\text{٣ م (السكريات ضد المراقبة)} = \frac{\text{٢٣٩٦}^2}{\text{٤٠}} + \frac{\text{٧٠١}^2}{\text{١٠}} - \frac{\text{٣٠٩٧}^2}{\text{٥٠}}$$

$$= \text{٨٣٢,٣٢} \quad \text{حيث} \quad \nu_2 = \text{٢} - \text{١} = \text{١}$$

$$\text{تقدير التباين بين القسمين} \quad = \text{ع}^2_{\text{ب}} = \text{٨٣٢,٣٢} \div \text{١} = \text{٨٣٢,٣٢}$$

$$\text{ف}_{\text{ى}} = \frac{\text{٨٣٢,٣٢}}{\text{٥,٤٦}} = \text{١٥٢,٤٤}$$

$$\text{من الجدول : ف}_{\text{٠,٠١} \, [\text{١ ، ٤٥}]} = \text{٧,٢٣}$$

نرفض الفرض الصفرى عند مستوى الدلالة ٠,٠١ ونظراً لأن متوسط السكريات يقل عن متوسط المراقبة نستنتج أن إضافة السكريات يؤخر نمو مقاطع نبات البسلة .

مثال (٨ – ٣) مقارنة السكريات النقية ضد السكر الخليط :

قارن تأثير إضافة السكريات النقية (مجتمعة) وتأثير إضافة السكر الخليط .

الحل :

الفرض الصفرى هو أن متوسط أقسام السكريات النقية معاً يساوى متوسط قسم السكريات الخليط .

نعتبر هنا أيضاً أن لدينا قسمين يتألف الأول من الأقسام ١ ، ٢ ، ٤ ككل وبها ٣٠ عنصراً مجموع قيمها ١٨١٦ ومتوسطها ٦٠,٥٣ ويتألف الثاني من القسم ٣ وبه ١٠ عناصر مجموع قيمها ٥٨٠ ومتوسطها ٥٨ ويكون المجموع الكلى لقيم العناصر التي اخترناها للدراسة ٢٣٩٦ .

مج ٢ (سكريات نقية ضد سكريات خليط) $= \frac{١٨١٦^2}{٣٠} + \frac{٥٨٠^2}{١٠} - \frac{٢٣٩٦^2}{٤٠}$

$=$ ٤٨,١٣ حيث $\nu_2 =$ ٢ − ١ = ١

$ف_ى = \frac{٤٨,١٣}{٥,٤٦} =$ ٨,٨٢

من الجدول : $ف_{٠,٠١\ (١ ،\ ٤٥)} =$ ٧,٢٣

نرفض الفرض الصفرى عند مستوى الدلالة ٠,٠١ ، ونظراً لأن متوسط مجموعة السكريات النقية أكبر من متوسط مجموعة السكر الخليط نستنتج أن إضافة السكريات النقية يؤخر نمو النبات بدرجة أقل مما يؤخره السكر الخليط .

مثال (٨ – ٤) مقارنة مجموعات السكريات النقية معاً :

ابحث ما إذا كان هناك فرق جوهرى بين تأثير السكريات النقية الثلاثة .

الحل :

الفرض الصفرى هو عدم وجود فروق بين متوسطات الأقسام الثلاثة . نعتبر هنا أن لدينا ثلاثة أقسام ١ ، ٢ ، ٤ مجموع قيمها ١٨١٦ .

مج ٢ (بين الأقسام) $= \frac{٥٩٣^2}{١٠} + \frac{٥٨٢^2}{١٠} + \frac{٦٤١^2}{١٠} - \frac{١٨١٦^2}{٣٠}$

$=$ ١٩٦,٨٧ حيث $\nu_1 =$ ٣ − ١ = ٢

تقدير التباين بين الأقسام $= ع^2_ب =$ ١٩٦,٨٧ ÷ ٢ = ٩٨,٤٣٥

$ف_ى = \frac{٩٨,٤٣٥}{٥,٤٦} =$ ١٨,٠٦

من الجدول نجد أن $ف_{٠,٠١ (٢ ، ٤٥)}$ تقع بين ٥,١٨ ، ٤,٩٨ وإذن نرفض الفرض الصفرى عند مستوى الدلالة ٠,٠١ ونستنتج أن السكريات النقية تختلف في تأثيرها ، ويبدو أن هذا الاختلاف يرجع إلى السكروز الذى له متوسط أعلى بكثير من متوسط النوعين الآخرين .

نستطيع تلخيص ما توصلنا إليه حتي الآن في الجدول الآتي :

جدول (٨ – ٥)

مصدر الاختلاف	مجموع المربعات	درجات الحرية	تقدير التباين	ف ى
بين الأقسام	١٠٧٧.٣٢	٤	٢٦٩.٣٣	٤٩.٣٣
المراقبة ضد السكريات	٨٣٢.٣٢	١	٨٣٢.٣٢	١٥٢.٤٤
السكريات النقية ضد الخليط	٤٨.١٣	١	٤٨.١٣	٨.٨٢
بين السكريات النقية	١٩٦.٨٧	٢	٩٨.٤٤	١٨.٠٣
داخل الأقسام	٢٤٥.٥٠	٤٥	٥,٤٦	
المجموع	١٣٢٢.٨٢	٤٩		

ملاحظة (٥) :

إن تحديد شكل وعدد الاختبارات القبلية يتوقف على التساؤلات التي تطرحها المشكلة . على أن هناك تحفظات ينبغى مراعاتها ، فلا يجب أن يزيد مجموع درجات الحرية للمقارنات القبلية عن ك – ١ حيث ك عدد الأقسام ، ومن الواضح إذن أنه من غير المناسب أن نقرر مقدماً إجراء المقارنة بين متوسطات كل زوج من الأقسام وهى تتطلب ½ ك (ك – ١) من المقارنات ، هذا العدد أكبر من ك – ١ حين ك > ٢ .

وبالإضافة إلى ذلك يفضل أن تختار الاختبارات القبلية بحيث تكون مستقلة ، وذلك لكى تكون المعلومات الناتجة من أى منها ذات قيمة بذاتها وغير متداخلة مع المعلومات الناتجة من أى مقارنة أخرى ، وهذا ما فعلنا في المثال السابق ، إذ أجرينا الاختبارات المستقلة الآتية :

(١) ضد (٢) ضد (٤) بدرجات حرية عددها ٢

(١ ، ٢ ، ٤) ضد (٣) بدرجات حرية عددها ١

(١ ، ٢ ، ٣ ، ٤) ضد (٥) بدرجات حرية عددها ١

وقد أدى هذا الاستقلال إلى أن يكون مجموع مجاميع المربعات في المقارنات الثلاث التي بنيت على ٢ ، ١ ، ١ من درجات الحرية مساوياً لمجموع المربعات بين الأقسام في التحليل العام الذى بني على ٤ درجات حرية . وهذا واضح في الجدول (٨ - ٥) . أى أن مجموع المربعات بين الأقسام قد تحلل إلى ثلاثة أجزاء منفصلة كل منها هو مجموع مربعات قائم بذاته وله درجات حرية خاصة به .

(٨ - ٥ - ٢) الاختبارات البعدية :

هى تلك الاختبارات التي لم يخطط لإجرائها أثناء تصميم التجربة ولكنها تقترح نفسها عند التأمل فيما وصلنا إليه من نتائج بعد إجراء التجربة وتحليل البيانات ، إذ أن هذا التأمل يجعلنا نشتبه في وجود فروق جوهرية بين بعض الأقسام مما يستحق البحث والاختبار .

ففي التجربة التي بالمثال (٨ - ١) نشعر بأن هناك فرقاً كبيراً بين متوسط السكروز والمتوسطات الأخرى من السكريات مما يوحي بضرورة اختبار دلالة هذا الفرق . كذلك نشعر بأن الفرق بين متوسط السكروز ومتوسط المراقبة يستحق الاختبار .

إن مثل هذه الاختبارات لا تجرى إلا إذا كانت النتيجة العامة لتحليل التباين

ذات دلالة أى حين يرفض الفرض الصفرى عن تساوى جميع المتوسطات لأنه لو ظهر أن هذه المتوسطات متساوية فإن هذا يتضمن أن الفروق الظاهرة بين أى قسمين لا تكون فروقاً ذات دلالة وبالتالى فإن أى اختبار نجريه لا يضيف جديداً لما علمناه عن دلالة هذه الفروق .

على أن المقارنات البعدية تحتاج لتقرير دلالتها إلى طرق خاصة تختلف عن تلك التي استخدمت في التحليل العام وفى المقارنات القبلية . ذلك لأن الأقسام التى نأخذها للمقارنة ولو أنها مأخوذة من نفس المجتمع العام إلا أننا نقتطعها عمداً من جزء متحيز من التوزيع فهى تفتقد عنصر العشوائية ولا يصبح توزيع الاحتمال الذى أسست عليه عملية اختبار الفروض صالحاً لها . وإذا تناولنا التجربة بالمثال (٨ — ١) وقارناً القسم (٣) الذى أعطى أصغر متوسط والقسم (٥) الذى أعطى أكبر متوسط نكون قد أخذنا الجزء المتطرف الأيسر والجزء المتطرف الأيمن من التوزيع ويكون من السهل ظهور اختلاف كبير بين متوسطيهما حتي ولو كانا من نفس المجتمع . وإذا أردنا الدقة في الحكم فيجب أن نأخذ هذا في الاعتبار وذلك بتصعيب تقرير دلالة مثل هذا الفرق .

ويعتمد أحد طرق الاختبارات البعدية على إيجاد قيمة حرجة لمجموع المربعات إذا تعداها مجموع المربعات بين قسمين أو مجموعة من الأقسام يكون الاختلاف بينها ذا دلالة وإلا فهو ليس كذلك . وتوجد هذه القيمة الحرجة كما يلي :

إن الاختبارات البعدية لا تجرى كما سبق القول إلا إذا كانت قيمة $\text{ف}_{\text{ى}}$ فى تحليل التباين ذات دلالة ، أى إذا كان :

$$\text{ف}_{\text{ى}} = \text{ع}^{٢}_{\text{ب}} / \text{ع}^{٢}_{\text{خ}} \geq \text{ف}_{\alpha}[\text{ك} - ١ ، \text{ن} - \text{ك}]$$

$$\text{أى} \quad \frac{\text{م م (بين الأقسام)}}{(\text{ك} - ١)\ \text{ع}^{٢}_{\text{خ}}} \geq \text{ف}_{\alpha}[\text{ك} - ١ ، \text{ن} - \text{ك}]$$

$$\text{أى} \quad \text{م م (بين الأقسام)} \geq (\text{ك} - ١)\ \text{ع}^{٢}_{\text{خ}}\ \text{ف}_{\alpha}[\text{ك} - ١ ، \text{ن} - \text{ك}] \quad (٧)$$

والعدد الذى بالطرف الأيسر من هذه المتباينة هو القيمة الحرجة المطلوبة . ويلاحظ أن هذا العدد أكبر من القيمة الحرجة المناظرة التي كان من الممكن استخدامها في المقارنات القبلية بعد وضع عدد أقسام المقارنة ا بدلا من العدد الكلى للأقسام ك . والهدف من كبر هذه القيمة تصعيب تقرير دلالة الاختلاف تعويضاً عن أننا نختار للمقارنة تلك الأقسام التي تسهم إسهاماً كبيراً في دلالة تحليل التباين .

ففي المثال (٨ – ١) وبأخذ α = ٠,٠٥ نجد أن :

القيمة الحرجة لمجموع المربعات = ٤ × ٥,٤٦ × ٢,٥٨ = ٥٦,٣٥

فإذا زادت قيمة مجموع المربعات بين متوسطات قسمين أو أكثر عن هذا العدد أو كانت مساوية له فإنها تكون ذات دلالة عند المستوى ٠,٠٥ .

ملاحظة (٦) :

تسمى هذه الطريقة « إجراء الاختبار الآنى لمجموع المربعات » sum of squares simultaneous test procedure وهى إحدى طرق الاختبارات البعدية للمقارنات المتعددة .

مثال (٨ – ٥) :

اختبر ما تراه يستحق الاختبار في نتائج تجربة المثال (٨ – ١) .

الحل :

لو رتبنا المتوسطات الناتجة تصاعدياً نجد الآتي :

٥٨ (جلوكوز + فركتوز) ، ٥٨,٢ (فركتوز) ، ٥٩,٣ (جلوكوز) ، ٦٤,١ (سكروز) ، ٧٠,١ (مراقبة) . وهذا الترتيب يوحي بعدة مقارنات نكتفي منها بما يلي :

(أولا) المقارنة بين المتوسطات الثلاث الأولى ، حيث أنها تبدو قريبة من بعضها ويشك في وجود فرق جوهرى بينها .

$$م\ م\ (\text{بين الأقسام الثلاثة}) = \frac{٥٩٣^{٢} + ٥٨٢^{٢} + ٥٨٠^{٢}}{١٠} - \frac{(٥٩٣ + ٥٨٢ + ٥٨٠)^{٢}}{٣٠}$$

$$= ١٠٢٦٧٧,٣ - ١٠٢٦٦٧,٥ = ٩,٨$$

بما أن ٩,٨ أصغر من القيمة الحرجة ٥٦,٣٥ نحكم بأن الفروق بين المتوسطات الثلاث ليست ذات دلالة أى يمكن اعتبار أن هذه الأقسام من مجتمعات متساوية المتوسطات ، وذلك عند مستوى الدلالة ٠,٠٥ .

(ثانياً) المقارنة بين متوسط السكروز ومتوسط السكريات الثلاثة الأخرى :
م م (سكروز ضد السكريات الأخرى) .

$$= \frac{٦٤١^{٢}}{١٠} + \frac{(٥٩٢ + ٥٨٢ + ٥٨٠)^{٢}}{٣٠} - \frac{(٦٤١ + ٥٩٣ + ٥٨٢ + ٥٨٠)^{٢}}{٤٠}$$

$$= ٤١٠٨٨,١ + ١٠٢٦٦٧,٥ - ١٤٣٥٢٠,٤ = ٢٣٥,٢$$

بما أن ٢٣٥,٢ أكبر من القيمة الحرجة ٥٦,٣٥ نحكم بأن الفرق بين السكروز والمستويات الأخرى من السكريات هو فرق جوهرى مما يشير إلى أن السكروز يؤخر نمو النبات بدرجة أقل من السكريات الأخرى .

ملاحظة (٧) :

إذا رغبنا في إجراء اختبار بين أى زوج من الأقسام فيمكننا استخدام اختبار ت للمقارنة بين متوسطى مجتمعين معتدلين – راجع البند (٦ – ٦ – ٣) مع أخذ $ع^{٢} = ع^{٢}_{خ}$ لتقدير التباين أى نستخدم الإحصاءة :

$$\frac{(\bar{س}_1 - \bar{س}_2) - (\mu_1 - \mu_2)}{ع_خ \sqrt{\frac{1}{ن_1} + \frac{1}{ن_2}}}$$

بدرجات حرية (ن – ك) وهي درجة الحرية لتقدير التباين داخل الأقسام في تحليل التباين ، إلا أن الطريقة سابقة الذكر أكثر عمومية ، فهى صالحة للتطبيق مهما كان عدد أقسام المقارنة .

كما يمكننا إيجاد حدى ثقة بدرجة (١ – α) للفرق بين المتوسطين من الصيغة .

$$(\bar{س}_1 - \bar{س}_2) \pm ع_خ \sqrt{\frac{1}{ن_1} + \frac{1}{ن_2}} \; ت_{\alpha\,(ن - ك)} \qquad (٨)$$

أما حدا الثقة بدرجة (١ – α) لمتوسط أى قسم $\bar{س}_ر$ فهما :

$$\bar{س}_ر \pm \frac{ع_خ}{\sqrt{ن_ر}} \; ت_{\alpha\,(ن - ك)} \qquad (٩)$$

ملاحظة (٨) : القيمة الحرجة للفرق بين متوسطى عينتين

CRITICAL DIFFERENCE (C.D.)

إذا رغبنا في إجراء عدة اختبارات بعدية بين أزواج من المتوسطات فيمكن بنفس فكرة القيمة الحرجة لمجموع المربعات إيجاد القيمة الحرجة للفرق بين أى زوج $\bar{س}_1$ ، $\bar{س}_2$ من المتوسطات كالآتي (على فرض تساوى عدد الوحدات في كل مجموعة) :

يكون الفرق بين متوسطين $\bar{س}_1$ ، $\bar{س}_2$ ذا دلالة

$$\text{إذا كان } \frac{\left|\bar{س}_1 - \bar{س}_2\right|}{ع_خ \sqrt{\frac{1}{ن_1} + \frac{1}{ن_2}}} > ت_{\alpha\,(ن - ك)}$$

$$\text{أى إذا كان } \left|\bar{س}_1 - \bar{س}_2\right| \geqslant ع_خ \sqrt{\frac{1}{ن_1} + \frac{1}{ن_2}}\; ت_{\alpha\,(ن - ك)}$$

والعدد الذى بالطرف الأيسر من هذه المتباينة هو القيمة الحرجة المطلوبة .

ففى المثال (٨ — ١) وبأخذ α = ٠,٠٥ نجد أن :

القيمة الحرجة بين متوسطى قسمين = $\sqrt{٥,٤٦} \times \sqrt{٠,٢} \times ٢,٠١$

= ٢,١٠٠٥

إذا زاد الفرق $\left|\bar{س}_1 - \bar{س}_2\right|$ عن العدد ٢,١٠٠٥ أو كان مساويا له فإن هذا الفرق يكون ذا دلالة عند المستوى ٠,٠٥ وإلا فلا دلالة له عند هذا المستوى .

فمثلا : الفرق بين متوسطى قسمى الجلوكوز والفركتوز = ٥٩,٣ — ٥٨,٢ = ١,١

وهذا الفرق أصغر من القيمة الحرجة ٢,١ فهو غير ذى دلالة عند المستوى ٠,٠٥ . أما الفرق بين متوسطى قسمى السكروز والمراقبة وهو ٧٠,١ — ٦٤,١ = ٦ فأكبر من القيمة الحرجة ٢,١ فهو فرق جوهرى . ونصل إلى نفس هذه الاستنتاجات إذا استخدمنا طريقة القيمة الحرجة لمجموع المربعات .

ملاحظة (٥) : التقسيم الأحادى

ONE—WAY CLASSIFICATION

إن التقسيم الناتج عن التجارب ذوات العامل الواحد التى نوقشت فى البنود

السابقة يندرج تحت ما يسمى بالتقسيم الأحادى أو التقسيم من ناحية واحدة . وهذا التقسيم يتناول حالتين :

(ا) الحالة السابق دراستها ومُثل لها بالمثال (٨ – ١) وتوابعه ، حيث يكون لدينا **مجتمع واحد** ونريد اختبار تأثير ك **من المعالجات** (مستويات عامل التقسيم) على متغير ما متعلق بهذا المجتمع . فى هذه الحالة نسحب من هذا المجتمع عينة عشوائية ثم نقسمها عشوائيا إلى ك من المجموعات غير المتداخلة ونعطى لكل منها معالجة مختلفة ، ثم ندرس الاستجابات فى هذه المجموعات .

(ب) الحالة التى يكون لدينا فيها ك **من المجتمعات** لكل منها خاصة متميزة ويراد دراسة تأثير هذه الخواص على متغير ما . هنا نسحب عينة عشوائية من كل مجتمع ونعطى لكل منها **نفس المعالجة** ثم ندرس الاستجابات . ومن أمثلة هذه الحالة المسائل ٢ ، ٣ ، ٥ ، ٦ من التمارين (٨ – ١) الآتية . ففى المسألة (٢) لدينا ٤ مجتمعات هى مجتمعات الأنواع الأربعة من الأرانب ، وفى المسألة (٣) لدينا ٣ مجتمعات هى المجتمع الذى تزرع فيه البذور فى أول مايو والمجتمع الذى تزرع فيه البذور فى ١٥ مايو ثم مجتمع ٢٩ مايو .

فى الحالة (ب) نفترض أن المجتمعات التى سحبت منها العينات هى مجتمعات معتدلة لها نفس التباين ، ونستخدم نفس النموذج الإحصائى ، وبهذا لا يختلف التحليل الاحصائى نظريا ولا حسابيا عن الحالة (ا) .

تمارين (٨ – ١)

(افتراضات العشوائية والاعتدالية متوفرة)

(١) الجدول الآتي يبين عينة من قيم شدة المقاومة لمعدن معين (بعد طرح ١٠٠ من كل منها) وقد حصلنا على هذه القيم من ٤ أشرطة من المعدن قيس كل منها

عند ٣ نقط مختلفة (الركن – الوسط – الحافة) . هل متوسط شدة المقاومة واحد عند جميع نقط الشريحة ؟

الركن	الوسط	الحافة
٣٧	٤٠	٤٢
٤٢	٣٩	٤٠
٢٨	١٧	٣٣
٣٧	٣٧	٤١

(٢) القيم الآتية هى أطوال أذناب نوع معين من يرقات القرادة في عينات من ٤ أنواع من الأرانب (مقيسة بالميكرون) . هل هناك فروق جوهرية بين متوسطات الأطوال في هذه الأنواع ؟

(١)	(٢)	(٣)	(٤)
٣٨٠	٣٥٠	٣٥٤	٣٧٦
٣٧٦	٣٥٦	٣٦٠	٣٤٤
٣٦٠	٣٥٨	٣٦٢	٣٤٢
٣٦٨	٣٧٦	٣٥٢	٣٧٢
٣٧٢	٣٣٨	٣٦٦	٣٧٤
٣٦٦	٣٤٢	٣٧٢	٣٦٠
٣٧٤	٣٦٦	٣٦٢	
٣٨٢	٣٥٠	٣٤٤	
	٣٤٤	٣٤٢	
	٣٦٤	٣٥٨	
		٣٥١	
		٣٤٨	
		٣٤٨	

اطرح ٣٠٠ أو أى عدد مناسب

(٣) هل تاريخ زراعة القطن يؤثر في وزن المحصول الناتج من البذور ؟

الآتي هي الأوزان بالكيلوجرامات الناتجة من ٤ حقول قسم كل منها إلى ٣ أحواض :

١ مايو	١٥ مايو	٢٩ مايو
٣,٣٥	٣,٨٦	١,٩٩
١,٤٩	٢,٧١	٢,٨٩
٢,٤٤	٢,١٨	١,٦٨
٢,٤٤	١,٩٥	٢,١٣

(٤) قسم ٢٨ أرنباً عشوائيا إلى ٤ أقسام متكافئة وأعطى لكل قسم معالجة ما (مثلا : دواء – نظام تغذية – بيئة ..) والجدول الآتي يعطى مستوى السكر في الدم الناتج من هذه المعالجات . هل هناك دليل على وجود فروق بين تأثيرات هذه المعالجات على مستوى السكر في الدم ؟

المعالجـات

(١)	(٢)	(٣)	(٤)
١٧	٣٧	٣٥	٩
١٦	٣٦	٢٢	٨
٢٨	٢١	٣٥	١٧
٤	١٣	٣٨	١٨
٢١	٤٥	٣١	١
صفر	٢٣	٣٤	٣٤
٢٣	١٣	٤٠	١٣

(٥) في تجربة صناعية كان أحد المهندسين مهتماً بكيفية تغير متوسط امتصاص الرطوبة بين خمس مجموعات مختلفة من الخرسانة ، واستخدم لذلك ٦ عينات لكل مجموعة تعرضت للرطوبة لمدة ٤٨ ساعة فجاءت البيانات كما يلي .

المجموعة (الوزن ٪)				
١	٢	٣	٤	٥
٥٥١	٥٩٥	٦٣٩	٤١٧	٥٦٣
٤٥٧	٥٨٠	٦١٥	٤٤٩	٦٣١
٤٥٠	٥٠٨	٥١١	٥١٧	٥٢٢
٧٣١	٥٨٣	٥٧٣	٤٣٨	٦١٣
٤٩٩	٦٣٣	٦٤٨	٤١٥	٥٥٦
٦٣٢	٥١٧	٦٧٧	٥٥٥	٦٧٩
اطرح ٤٥٠ أو ٥٠٠ أو أي عدد مناسب .				

اختبر أن $\mu_1 = \mu_2 = \cdots = \mu_5$ عند المستوى ٠,٠٥

(٦) في تجربة ما اختيرت سلالتان من Drosophila Melanogaster إحداهما لدورة يرقات قصيرة (short larval period) والأخرى لدورة طويلة ، كما أخذت سلالة كمجموعة مراقبة . وفي الجيل ٤٢ لخصت البيانات عن طول الدورة بالساعات فيما يلي :

	السلالة		
	دورة قصيرة	دورة طويلة	مراقبة
$ن_ي$:	٨٠	٣٣	٦٩
$ك_ي$:	٨٠٧٠	٣٦٤٠	٧٢٩١
مجـ مجـ $س^2_{ري}$:	١٩٩٤٦٥٠		

أولا : أجر تحليل التباين وفسره .
ثانيا : أجر المقارنة القبلية بين الدورتين القصيرة والطويلة معاً ضد المراقبة .
ثالثا : أجر المقارنة البعدية بين كل من أزواج المتوسطات الثلاثة .
رابعاً : أوجد فترة ثقة بدرجة ٩٥٪ لكل من هذه المتوسطات .

(٧) في أحد الأبحاث الطبية عن علاج جحوظ العين الناتج من تسمم الغدة الدرقية اختيرت عينة عشوائية من ٣٠ ضفدعة وقسمت عشوائياً إلى ثلاث مجموعات متساوية العدد تلقت إحداها علاجاً دوائياً وتلقت أخرى علاجاً جراحياً (استئصال الغدة النخامية) وتركت المجموعة الثالثة دون علاج (مجموعة مراقبة) ثم قيست العيون بالمقياس الأفقي للعين فنتجت القيم الآتية بالملليمترات .

مجموعة العلاج الدوائي : ١,٦ ١,٧ ١,٤ ١,٨ ١,٩ ١,٨ ١,٣ ١,٦ ١,٥ ١,٧
مجموعة العلاج الجراحي : ١,٥ ١,٤ ١,٣ ١,٢ ١,٨ ١,٩ ١,٩ ١,٧ ١,٦ ١,٣
مجموعة المراقبة : ١,٩ ٢ ٢ ١,٨ ٢ ١,٩ ٢,١ ٢ ١,٧ ١,٨

(أ) اختبر تأثير مجموعتي العلاج ضد مجموعة المراقبة .
(ب) اختبر ما إذا كان هناك فرق جوهرى بين نوعى العلاج .

(خذ α = ٠,٠١) .

(٨ – ٦) التجارب ذوات العاملين :

TWO—FACTOR EXPERIMENTS

اعتبر عينة حجمها ن مأخوذة من متغير معتدل س وسطه الحسابي μ وتباينه σ^2 وافرض أن وحدات هذه العينة قد تعرضت لتأثير عاملين لأحدهما ك من المستويات وللثاني هـ من المستويات . المطلوب اختبار ما إذا كان المجتمع الذى أخذت منه العينة متجانساً بالنسبة لكل من هذين العاملين على حدة .

في هذه الحال علينا أن نختار بين طريقتين للتحليل ، ويتوقف هذا الاختيار على ما إذا كان العاملان مستقلين أو كانا يتفاعلان معاً . والمقصود بكلمة التفاعل هنا هو تأثير أى من العاملين على الآخر . وسنبدأ بالحالة الأبسط التي سنفترض فيها عدم وجود تفاعل بين العاملين .

(٨ – ٦ – ١) حالة عاملين لا يتفاعلان :

على أساس توفر عوامل الاعتدالية والعشوائية وخطية النموذج وعلى فرض عدم وجود تفاعل بين عاملى التجريب ، يكون تحليل التباين ما هو إلا امتداد لتحليل التباين للتجارب ذوات العامل الواحد ولا يزيد عنه إلا بخطوة منطقية واحدة . توضع وحدات العينة في ك من الأعمدة تناظر مستويات العامل الأول ، هـ من الصفوف تناظر مستويات العامل الثاني ، فتخضع أى وحدة تجريبية $س_{رو}$ للنموذج الآتي :

$$س_{رو} = \mu + \alpha_{و} + \beta_{ر} + خ_{رو} \qquad (١٠)$$

حيث μ ، $\alpha_{و}$ ، $خ_{رو}$ تحمل نفس المعاني السابقة في النموذج (٢) ، وحيث $\beta_{ر}$ تعبر عن متوسط أثر المستوى ر للعامل الثاني .

وكنتيجة مباشرة لذلك ، لا يتحلل مجموع المربعات الكلى إلى مركبتين كما هو الحال في المتساوية (٣) بل إلى ثلاث مركبات . ومن الطبيعى أن تظل المركبة الأولى التي تعبر عن الاختلاف بين مستويات العامل الأول (الأعمدة) كما هى ، أما المركبة الثانية في المتساوية (٣) فتنفصل إلى مركبتين إحداهما تعبر عن الاختلاف بين مستويات العامل الثاني (الصفوف) والأخرى تعطى الاختلاف الذى يتبقى من الاختلاف الكلى بعد استبعاد أثر كل من العاملين . هذا الاختلاف يعزى إلى عدة أسباب نضمها تحت كلمة « الخطأ » ، منها الخطأ العشوائي أو خطأ التجريب ومنها الخطأ الناشيء عن إهمال التفاعل بين العاملين إذا كان هناك تفاعل .

يمكن جبرياً أن نثبت المتطابقة الآتية :

$$\sum_{ر}^{ك}\sum_{و}^{هـ}(س_{ر و}-\bar{س})^{٢}=\sum_{و}ن_{و}(\bar{س}_{و}-\bar{س})^{٢}+\sum_{ر}ن_{ر}(\bar{س}_{ر}-\bar{س})^{٢}$$

$$+\sum_{و}\sum_{ر}(س_{ر و}-\bar{س}_{و}-\bar{س}_{ر}+\bar{س})^{٢} \qquad (٩)$$

وهذه المتطابقة تترجم لفظيا كالآتى :

م م (الكلى) = م م (بين الأعمدة) + م م (بين الصفوف)
+ م مَ (الخطأ)

حيث $\nu_{١} = ن - ١$ ، $\nu_{٢} =$ ك $- ١$ ، $\nu_{٣} =$ هـ $- ١$

، $\nu_{٤} = (ن - ك) - ($هـ$ - ١) = ($ك$ - ١) ($هـ$ - ١)$ (١٠)

وإذن :

$$\nu_{١} = \nu_{٢} + \nu_{٣} + \nu_{٤}$$

إن هذه المركبات الثلاث تعطى ثلاثة تقديرات مستقلة لتباين المجتمع $\sigma^{٢}$ وهى :

$ع_{١}^{٢}$ = مم (بين الأعمدة)/(ك−١) ، $ع_{٢}^{٢}$ = مم (بين الصفوف)/(هـ − ١) ،

$ع_{٣}^{٢}$ = مم (الخطأ) / (ن − ك − هـ + ١)

ولا يبقى إلا استخدام نسبة التباين $\frac{ع_{١}^{٢}}{ع_{٣}^{٢}}$ لاختبار صحة الفرض الصفرى عن

تساوى متوسطات مستويات العامل الأول . واستخدام نسبة التباين $\frac{ع_{٢}^{٢}}{ع_{٣}^{٢}}$ لاختبار

صحة الفرض الصفرى عن تساوى متوسطات مستويات العامل الثاني .

مثال (٨ – ٦) :

قسمت ١٢ بقرة إلى هـ = ٤ من المجموعات بكل منها ٣ بقرات بحسب الوزن عند بدء التجربة . أعطى للأبقار الثلاث بكل مجموعة نوع مختلف من الغذاء . وبعد فترة من الزمن قيست الزيادات في أوزان الأبقار ورصدت بالجدول (٨ – ٦) الآتى . لدينا عاملان الأول هو عامل الغذاء وله ك = ٣ مستويات والثانى هو عامل الوزن الابتدائى للأبقار وله هـ = ٤ مستويات . المطلوب بحث تأثير كل من هذين العاملين على الزيادة فى وزن الأبقار عند مستوى الدلالة ٠,٠٥

جدول (٨ – ٦)

الوزن الابتدائى للأبقار	نوع الغذاء أ١	أ٢	أ٣	نر	مر
و١	٧,٠	١٤,٠	٨,٥	٣	٢٩,٥
و٢	١٦,٠	١٥,٥	١٦,٥	٣	٤٨,٠
و٣	١٠,٥	١٥,٠	٩,٥	٣	٣٥,٠
و٤	١٣,٥	٢١,٠	١٣,٥	٣	٤٨,٠
نن	٤	٤	٤	ن = ١٢	
من	٤٧,٠	٦٥,٥	٤٨,٠	م = ١٦٠,٥	

الحل :

لدينا اثنان من الفروض الصفرية : لا الوزن الابتدائى ولا نوع الغذاء له تأثير فى زيادة أوزان الأبقار .

مم (الكلى) = ٧٢ + ... + ١٤٢ + ... ٨,٥٢ ... + ١٣,٥٢ − $\frac{١٦٠,٥^{٢}}{١٢}$

= ٢٣١٦,٧٥ − ٢١٤٦,٦٨٧٥

= ١٧٠,٠٦٢٥   بدرجات حرية ١١

مم (بين الأعمدة) = $\frac{٤٧^{٢} + ٦٥,٥^{٢} + ٤٨^{٢}}{٤}$ − $\frac{١٦٠,٥^{٢}}{١٢}$

= ٢٢٠٠,٨١٢٥ − ٢١٤٦,٦٨٧٥

= ٥٤,١٢٥   بدرجات حرية ٢

مم (بين الصفوف) = $\frac{٢٩,٥^{٢} + ٤٨^{٢} + ٣٥^{٢} + ٤٨^{٢}}{٣}$ − $\frac{١٦٠,٥^{٢}}{١٢}$

= ٢٢٣٤,٤١٦٦ − ٢١٤٦,٦٨٧٥

= ٨٧,٧٢٩١   بدرجات حرية ٣

مم (الخطأ) = ١٧٠,٠٦٢٥ − (٥٤,١٢٥ + ٨٧,٧٢٩١)

= ٢٨,٢٠٨٤   بدرجات حرية ٦

وينشأ جدول التباين الآتى .

الجدول (٨ − ٧)

مصدر التباين	مجموع المربعات	درجات الحرية	تقدير التباين	ف
بين الأعمدة (نوع الغذاء)	٥٤,١٢٥٠	ك − ١ = ٢	$ع_{١}^{٢}$ = ٢٧,٠٦٢٥	٥,٧٥٦*
بين الصفوف (الوزن الابتدائى)	٨٧,٧٢٩١	هـ − ١ = ٣	$ع_{٢}^{٢}$ = ٢٩,٢٤٣٠	٦,٢٢٠*
خطأ التجريب	٢٨,٢٠٨٤	(ك − ١)(هـ − ١) = ٦	$ع_{خ}^{٢}$ = ٤,٧٠١٤	
المجموع	١٧٠,٠٦٢٥	ن − ١ = ١١		

من الجدول نجد أن ف$_{٠,٠٥[٢، ٦]}$ = ٥,١٤ وهذه أصغر من ٥,٧٥٦ وإذن نرفض الفرض الصفرى الأول عند المستوى ٠٥,، ونقرر أن اختلاف نوع الغذاء يؤدى إلى الاختلاف فى الزيادة فى أوزان الأبقار .

كذلك ف$_{٠,٠٥[٣، ٦]}$ = ٤,٧٦ وهذه أصغر من ٦,٢٢٠ واذن نرفض الفرض الصفرى الثانى عند المستوى ٠,٠٥ ونقرر أن اختلاف الأوزان الابتدائية للأبقار له تأثير فى الزيادة فى أوزانها .

تمارين (٨ – ٢)

(١) أجريت تجربة زراعية لاختبار تأثير اختلاف التربة (٥ قطع من الأرض) واختلاف نوع القمح (٧ سلالات) على محصول الحبوب . وقد قسمت كل قطعة أرض عشوائياً إلى ٧ أحواض وزعت عليها السلالات السبع عشوائياً فنتجت البيانات الآتية التي تسجل المقادير الناتجة للمحصول بالكيلة .

قطعة الأرض	السلالة (١)	(٢)	(٣)	(٤)	(٥)	(٦)	(٧)	$م_س$
(١)	١٦	١٥	١٧	١٤	١٤	١٥	١٣	١٠٤
(٢)	١٢	٩	١٥	١٠	١٠	١١	١١	٧٨
(٣)	١٣	١٣	١٤	١٥	١٢	١٣	١٠	٩٠
(٤)	١٥	١٤	١٩	١٧	١٣	١٨	١٦	١١٢
(٥)	١١	١٠	١٢	١٠	١١	١٢	١٢	٧٨
$م_ي$	٦٧	٦١	٧٧	٦٦	٦٠	٦٩	٦٢	٤٦٢

مج مج س^{٢} = ٦٣١٤

أولا : ابحث دلالة تأثير كل من عاملى التربة ونوع القمح .

ثانيا : ابحث ما إذا كان الفرق بين السلالتين (٥) ، (٦) ذا دلالة عند المستوى ٠,٠٥

ثالثا : ابحث ما إذا كانت تربة القطعتين ٢ ، ٥ (معاً) تختلف عن تربة القطعتين ١ ، ٤ (معاً) .

(استخدام مستوى الدلالة ٠,٠١) .

(٢) الآتي هى مقادير الكلوسترول (بالملليجرام في العبوة) التي وجدتها ٤ معامل في عبوات لثلاثة أنواع متشابهة من الغذاء وزن كل منها ٦ أوقيات .

الغذاء	المعامل			
	(١)	(٢)	(٣)	(٤)
(أ)	٣,٧	٢,٨	٣,١	٣,٤
(ب)	٣,١	٢,٦	٢,٧	٣,٠
(حـ)	٣,٥	٣,٤	٣,٠	٣,٣

اختبر عند مستوى الدلالة ٠,٠٥ ما إذا كانت :

(أولا) متوسطات الكلوسترول في الأنواع الثلاثة من الغذاء متساوية .

(ثانياً) المتوسطات التى حصلت عليها المعامل الأربعة متساوية .

(٣) الأعداد الآتية هى درجات الحرارة (مقاسة بالسنتجراد) لمياه إحدى البحيرات في أربعة أيام متتالية من صيف ١٩٥٢ م (الساعة الثانية بعد الظهر) ، وقد أخذت هذه الدرجات على ١٠ أعماق مختلفة (مقاسة بالمتر) . ابحث دلالة كل من عاملى اليوم والعمق .

الأعماق	٢٩ يوليو	٣٠ يوليو	٣١ يوليو	١ أغسطس	$م_{ر}$
٠	٢٣,٨	٢٤,٠	٢٤,٦	٢٤,٨	٩٧,٢
١	٢٢,٦	٢٢,٤	٢٢,٩	٢٣,٢	٩١,١
٢	٢٢,٢	٢٢,١	٢٢,١	٢٢,٢	٨٨,٦
٣	٢١,٢	٢١,٨	٢١,٠	٢١,٢	٨٥,٢
٤	١٨,٤	١٩,٣	١٩,٠	١٨,٨	٧٥,٥
٥	١٣,٥	١٤,٤	١٤,٢	١٣,٨	٥٥,٩
٦	٩,٨	٩,٩	١٠,٤	٩,٦	٣٩,٧
٩	٦,٠	٦,٠	٦,٣	٦,٣	٢٤,٦
١٢,٥	٥,٨	٥,٩	٦,٠	٥,٨	٢٣,٥
١٥,٥	٥,٦	٥,٦	٥,٥	٥,٦	٢٢,٣
$م_{ي}$	١٤٨,٩	١٥١,٤	١٥٢,٠	١٥١,٣	٦٠٣,٦

$مجـ\ مجـ\ س^{٢}_{ري} = ١١٢٣٠,٧٨$

(٨ – ٦ – ٢) حالة عاملين يتفاعلان :

في التجارب ذوات العاملين ، إذا كان هناك شك في وجود تفاعل بين العاملين أى في تأثر كل منهما جزئياً بالآخر ، فإن النموذج (١٠) لا يكون مناسباً لأنه لا يفسح مكاناً لتأثير هذا التفاعل . وعلى فرض توفر شروط الاعتدالية والعشوائية وخطية النموذج فإن النموذج المناسب يأخذ الصورة الآتية :

$$س_{ري} = \mu + \alpha_{ي} + \beta_{ر} + (\beta\alpha)_{ري} + خ_{ري} \qquad (١١)$$

حيث $(\beta\alpha)_{رو}$ تعبر عن تأثير الاختلاف الناشيء من تفاعل العاملين . ولكى نوجد مقياساً لهذا الاختلاف لا مفر من تكرير التجربة برمتها مرة واحدة على الأقل وذلك لأنه في التجربة الواحدة يؤثر المستوى ٣ مثلا من العامل الأول مع المستوى ٢ مثلا من العامل الثاني على وحدة واحدة فقط هى $س_{٣٢}$ من وحدات التجريب ، وبالمثل بالنسبة لأزواج المستويات الأخرى ، ولا يكون هناك مجال حينئذ لإيجاد الاختلاف الناشيء عن تفاعل العاملين إلا إذا كان هناك أكثر من وحدة تتعرض لتأثير كل من أزواج هذه المستويات ، أى إلا إذا كررت التجربة مرة واحدة على الأقل . وتجمع نتيجة التجربيتن أو التجارب فيما يسمى بالخلايا كما في المثال (٨ — ٧) الآتي .

ومن الناحية الحسابية نوجد كلا من م٢ (الكلى) ، م٢ (بين الأعمدة) ، م٢ (بين الصفوف) من واقع القيم الناتجة عن التجريب كما في البند السابق . أما بالنسبة للاختلافين الباقيين فنحسبهما من مجاميع الخلايا كالآتي :

نفرض أننا كررنا التجربة برمتها أ من المرات فيكون كل زوج من أزواج المستويات قد أثر في أ من وحدات التجريب . سنسمى مجموع قيم كل من هذه الوحدات « **مجموع الخلية** » ونرمز له بالرمز $م_{س}$ وإذن :

$$م^{٢} \text{(بين الخلايا)} = \frac{مجـ\, م^{٢}_{س}}{أ} - \frac{م^{٢}}{ن} \qquad (١٢)$$

بدرجات حرية عددها (ك هـ — ١) حيث ك عدد الأعمدة ، هـ عدد الصفوف . إن هذا الاختلاف هو مجموع الاختلافات الناشئة عن العاملين بكل أنواعها وإذن :

م٢ (تفاعل العاملين) = م٢ (بين الخلايا) — م٢ (بين الأعمدة) — م٢ (بين الصفوف) بدرجات حرية عددها

(ك هـ —١) — (ك —١) — (هـ —١) = ك هـ — ك — هـ +١)

= (ك —١) (هـ —١)

أما الاختلاف الذى نضعه تحت كلمة « خطأ » فهو الاختلاف المتبقي من الاختلاف الكلى بعد استبعاد جملة الاختلاف الناشيء من العاملين . أى أن :

م م (الخطأ) = م م (الكلى) − م م (بين الخلايا) .

بدرجات حرية عددها (ن − ١) − (ك هـ − ١) = ن − ك هـ

مثال (٨ − ٧) :

الجدول (٨ − ٨) الآتي يعطى نتائج تجربة أجريت مرتين (بشكل مستقل) لدراسة تأثير كل من عاملى شدة الضوء ودرجة الحرارة على معدل نمو أحد النباتات . الأعداد ١٧ ، ٦ ، ٣ ، ٢ ، ... ، ٨ ، ٤ ، ٥ ، ٢ هى نتائج التجربة الأولى والأعداد ٩ ، ٨ ، ٤ ، ٢ ، ... ، ١٠ ، ٥ ، ٤ ، ٢ هى نتائج التجربة الثانية . أما الأعداد التى بين قوسين فهى مجاميع الخلايا . أجر تحليل التباين ثم أجب عن التساؤلات القبلية الآتية :

(أولا) هل متوسط تأثير درجة الحرارة ٨ يساوى متوسط تأثير درجة الحرارة ٢٨ ؟

(ثانيا) هل متوسط تأثير درجة الحرارة ١٤ يساوى متوسط تأثير درجة الحرارة ٢١ ؟

(ثالثا) هل متوسط تأثير درجتى الحرارة ٨ ، ٢٨ معا يساوى متوسط تأثير درجتى الحرارة ١٤ ، ٢١ معا ؟

الحل :

م م (الكلى) = مج مج مج س² − م² / ن

= ١٧² + ٩² + ٦² + ٨² + ... + ٥² + ٤² + ٢² + ٢² − ١٩٤² / ٣٢

= ١٧٧٤ − ١١٧٦ = ٥٩٨ حيث ν_1 = ٣٢ − ١ = ٣١

الجدول (٨ – ٨)

درجة الحرارة	شدة الضوء (١)	(٢)	(٣)	(٤)	$ن_س$	$م_س$
٨	١٧ ، ٩ (٢٦)	١٣ ، ١٥ (٢٨)	١٣ ، ١٦ (٢٩)	٨ ، ١٠ (١٨)	٨	١٠١
١٤	٦ ، ٨ (١٤)	٦ ، ٥ (١١)	٦ ، ٤ (١٠)	٤ ، ٥ (٩)	٨	٤٤
٢١	٣ ، ٤ (٧)	٣ ، ٥ (٨)	٣ ، ٤ (٧)	٥ ، ٤ (٩)	٨	٣١
٢٨	٢ ، ٢ (٤)	٢ ، ٣ (٥)	٢ ، ٣ (٥)	٢ ، ٢ (٤)	٨	١٨
$ن_ع$	٨	٨	٨	٨	ن = ٣٢	
$م_ع$	٥١	٥٢	٥١	٤٠	م = ١٩٤	

$$\text{م م (بين الأعمدة)} = \text{مج} \frac{م_ع^2}{ن_ع} - \frac{م^2}{ن}$$

$$= \frac{٥١^2 + ٥٢^2 + ٥١^2 + ٤٠^2}{٨} - ١١٧٦$$

$$= ١١٨٨,٢٥ - ١١٧٦$$

$$= ١٢,٢٥$$ حيث $\nu_2 = ٣$

$$\text{م م (بين الصفوف)} = \text{مج} \frac{م_س^2}{ن_س} - \frac{م^2}{ن}$$

$$= \frac{١٠١^2 + ٤٤^2 + ٣١^2 + ١٨^2}{٨} - ١١٧٦$$

$$= ١٦٧٧,٧٥ - ١١٧٦ = ٥٠١,٧٥$$ حيث $\nu_2 = ٣$

م م (بين الخلايا) $= \frac{\text{مجـ } م^{٢}_{س}}{أ} - \frac{م^{٢}}{ن}$

$= \frac{٢٦^{٢} + ١٤^{٢} + ... + ٩^{٢} + ٤^{٢}}{٢} - ١١٧٦$

= ١٧٢٤ − ١١٧٦

= ٥٤٨ حيث $\nu_{٤}$ = ٤×٤−١=١٥

م م (تفاعل) = ٥٤٨ − (١٢,٢٥ + ٥٠١,٧٥)

= ٣٤ حيث $\nu_{٥}$ = (٤−١)(٤−١) = ٩

، م م (الخطأ) = ٥٩٨ − ٥٤٨

= ٥٠ حيث $\nu_{٦}$ = ٣١−١٥=١٦

جدول (٨ – ٩)

مصدر التباين	مجموع المربعات	درجات الحرية	تقدير التباين	ف
بين الأعمدة (شدة الإضاءة)	١٢,٢٥	٣	٤,٠٨	١,٣٠٤
بين الصفوف (درجة الحرارة)	٥٠١,٧٥	٣	١٦٧,٢٥	٥٣,٤٣٥
تفاعل (الإضاءة × الحرارة)	٣٤	٩	٣,٧٧	١,٢٠٤
داخل الأقسام (الخطأ)	٥٠	١٦	٣,١٣	
المجموع	٥٩٨	٣١		

$ف_{٠,٠٥ [٣ ، ١٦]}$ تقع بين ٣,١٠ ، ٣,٢٩ ، $ف_{٠,٠١ [٣ ، ١٦]}$ تقع بين ٤,٩٤ ، ٥,٤٢

$ف_{٠,٠٥ [٩ ، ١٦]}$ تقع بين ٣٩ ، ٢,٥٩

وعلى ذلك فإن :

(١) الاختلافات بين مستويات شدة الإضاءة ليست ذات دلالة عند المستوى ٠,٠٥ (أى يمكن اعتبار هذه المستويات متكافئة) .

(٢) الاختلافات بين مستويات درجة الحرارة ذات دلالة عالية .

(٣) التفاعل بين عاملى الإضاءة ودرجة الحرارة ليس ذى دلالة عند المستوى ٠,٠٥

نجيب الآن عن التساؤلات الثلاثة المطروحة .

(أولا)

$$\text{مج}^2 \text{ (درجة الحرارة ٨ ضد درجة الحرارة ٢٨)} = \frac{١٠١^2}{٨} + \frac{١٨^2}{٨} - \frac{١١٩^2}{١٦} = ٤٣٠,٥٦٢٥$$

$$\text{ف}_{ى} = \frac{٤٣٠,٥٦٢٥}{٣,١٣} = ١٣٧,٥٦^{**}$$

بما أن $\text{ف}_{٠,٠١\,[١،١٦]} = ٨,٦٨ < ١٣٧,٥٦$ نرفض الفرض الصفرى عن تساوى تأثير درجة الحرارة ٨ وتأثير درجة الحرارة ٢٨ ونستنتج أن لدرجة الحرارة ٨ تأثير أكبر ، وذلك عند مستوى الدلالة ٠,٠١

(ثانيا)

$$\text{مج}^2 \text{ (بين مستوى درجتى الحرارة ١٤ ، ٢١)} = \frac{٤٤^2}{٨} + \frac{٣١^2}{٨} - \frac{٧٥^2}{١٦} = ١٠,٥٦٢٥$$

$$\text{ف}_{ى} = \frac{١٠,٥٦٢٥}{٣,١٣} = ٣,٣٧٥$$

بما أن $\text{ف}_{٠,٠٥\,[١،١٦]} = ٤,٥٤ > ٣,٣٧٥$ نقبل الفرض الصفرى عن تساوى تأثير درجتى الحرارة ١٤ ، ٢١ .

(ثالثا)

$$\text{م}^{٢}\text{(بين الزوجين)} = \frac{(١٠١ + ١٨)^{٢}}{١٦} + \frac{(٤٤ + ٣١)^{٢}}{١٦} - \frac{١٩٤^{٢}}{٣٢} = ٦٠,٥$$

$$\text{ف}_{\text{ى}} = \frac{٦٠,٥}{٣,١٣} = ١٩,٣٢٩^{**}$$

نرفض الفرض الصفرى عند مستوى الدلالة ٠,٠١

ملاحظة (٩) :

إن هذه المقارنات الثلاث مستقلة ويمكن التأكد من ذلك باستخدام معيار استقلال مقارنتين الذى سنقدمه بالبند (٨ – ١١ – ١) . ولذلك فإن مجموع مجاميع المربعات لهذه المقارنات وهو ٤٣٠,٥٦٢٥ + ١٠,٥٦٢٥ + ٦٠,٥ = ٥٠١,٦٢٥ يساوى مجموع المربعات بين الصفوف (درجات الحرارة) ، أى أن هذا المجموع قد تحلل إلى ٣ مركبات مستقلة ، كما تحللت درجات حريته الثلاثة إلى ثلاث درجات حرية منفصلة .

ملاحظة (١٠) :

إذا كنا قد وجدنا تفاعلا ذا دلالة بين شـدة الإضاءة ودرجة الحرارة فإن الخطوة التالية تكون حينئذ تقسيم البيانات إلى جداول منفصلة لكل مستوى من مستويات شدة الإضاءة وتحليل كل جدول لاختبار تأثير مستويات درجة الحرارة ، وقد نجد أنه في مستوى ما من مستويات الإضاءة تكون درجة الحرارة ذات تأثير جوهرى بينما في مستوى آخر لا يكون هذا التأثير جوهرياً . وبالمثل يمكن تجزىء البيانات لكل من مستويات درجة الحرارة لاختبار تأثير مستوى الإضاءة عند كل درجة حرارة .

تمارين (٨ – ٣)

في دراسة استهلاك الأكسجين لسلالتين من الحيوانات الصدفية عند ثلاثة تركيزات لماء البحر أخذت ٤ قراءات لكل مركب من السلالة وتركيز الماء (الملوحة) وسجلت القراءات بمقياس معين في الجدول الآتي :

الملوحة	السلالة (١)	السلالة (٢)	المجموع
٪١٠٠	٧,١٦	٦,١٤	
	٦,٧٨	٣,٨٦	
	١٣,٦٠	١٠,٤٠	
	٨,٩٤	٥,٤٩	
	مج = ٣٦,٤٧	مج = ٢٥,٨٩	٦٢,٣٦
٪٧٥	٥,٢٠	٤,٤٧	
	٥,٢٠	٩,٩٠	
	٧,١٨	٥,٧٥	
	٦,٣٧	١١,٨٠	
	مج = ٢٣,٩٥	مج = ٣١,٩٢	٥٥,٨٧
٪٥٠	١١,١١	٩,٦٣	
	٩,٧٤	٦,٣٨	
	٢٨,٨٠	١٣,٤٠	
	٩,٧٤	١٤,٥٠	
	مج = ٤٩,٣٩	مج = ٤٣,٩١	٩٣,٣٠
المجموع	١٠٩,٨١	١٠١,٧٢	٢١١,٥٣

(١) اختبر الفرض أن استهلاك الأوكسجين لا يختلف باختلاف السلالة .
(٢) اختبر الفرض أن استهلاك الأوكسجين لا يختلف باختلاف الملوحة .
(٣) اختبر تأثير تفاعل السلالة والملوحة على استهلاك الأوكسجين .

(خذ α = ٠,٠٥)

(٨ – ٧) المقارنات التزاوجية : PAIRED COMPARISONS

المفروض في التجارب ذوات العامل الواحد أن تكون وحدات التجريب فى مختلف أقسام المعالجة متجانسة بقدر الإمكان ، وذلك لكى تكون الفروق في استجابات هذه الوحدات لمستويات العامل راجعة فقط للفروق بين هذه المستويات . أما إذا كانت الوحدات غير متجانسة فإن هذا يتيح الفرصة لاختلافات عشوائية قد تكون من الكبر بحيث تؤدى إلى اختفاء الفروق الحقيقية في تأثيرات تلك المستويات كما سنرى في المثال (٨ – ٨) القادم .

غير أن متطلب التجانس هذا قد يفرض قيوداً عنيفة على عدد الوحدات التي تحتاجها الدراسة ، فمثلا لمقارنة نوعين من المسكنات لمرض ما المفروض أن يكون المرضي من نفس الجنس ومن نفس العمر والظروف الصحية وشدة الألم الناتج عن المرض ... ومن الواضح أنه من الصعب عمليا الحصول على عدد كاف من المرضي الذين يشتركون في هذه الخواص . وحتي إذا أمكن الحصول على مجموعة بهذه الأوصاف فإن دراسة مثل هذه المجموعة الخاصة تكون دراسة ضيقة وتكون النتائج قاصرة على مجموعة المرضي الذين يتصفون بهذه الصفات الخاصة ، والأفضل أن تشمل الدراسة مجالا أوسع لكى تكون النتائج أكثر عمومية ، وذلك بتجريب نوعى المسكنات على مختلف المرضي بذلك المرض تأخذهم عمداً من الجنسين ومن مختلف المجموعات العمرية والظروف الصحية .

ومن الضرورى إذا إيجاد حل وسط بين متطلب التجانس في وحدات التجريب ومتطلب تنويع هذه الوحدات ، وهما متطلبان متعارضان .

وقد وجد الحل بتصميم تجارب تعرف بتجارب القطاعات كاملة التعشية randomized complete blocks (كلمة كاملة تعني أن كل مستوى من مستويات العامل يظهر نفس العدد من المرات في كل قطاع) . وسنهتم هنا بحالة خاصة من هذه التجارب تعرف بالمقارنات التزاوجية وهى الحالة التي يكون فيها عامل التجريب ذا مستويين فقط (ك = ٢) وسميت تزاوجية لأن كل وحدة من وحدات التجريب تتلقي أحد المستويين ، تتزاوج مع وحدة **مماثلة** تتلقي المستوى الآخر . على أن الأسلوب الذى نتبعه في تناول وتحليل هذه الحالة يمتد إلى الحالات التي يكون فيها لعامل التجريب أكثر من مستويين .

والطريقة الإجرائية لهذه التجارب تعتمد على وضع وحدات التجريب مثني مثني في قطاعات يتألف كل منها من وحدتين تتلقي إحداهما المستوى الأول وتتلقي الأخرى المستوى الثاني ، بشرط أن تكون الوحدتان في كل قطاع متشابهتين – بل قد تكون نفس الوحدة تعالج مرتين في وقتين مختلفين – وبشرط أن يكون توزيع المستويين على الوحدتين عشوائياً (مثلا بإلقاء قطعة من العملة) في كل قطاع على حدة . أما الوحدات في القطاعات المختلفة فقد تكون متشابهة أو غير متشابهة .

القطاعات

	(١)	(٢)	(٣)		(ن)
وحدات	٢	١	١	....	٢
التجريب	١	٢	٢		١

إن هذا الإجراء يصون فعالية المقارنة داخل كل قطاع ويسمح في الوقت ذاته بتعدد الظروف بين مختلف القطاعات ، ويمكن النظر إلى الاستجابة في أي وحدة تجريبية على أنها مؤلفة من ثلاثة عناصر هي :

(أ) تأثير مستوى المعالجة لعامل التجريب .
(ب) تأثير الظروف داخل القطاعات (ويعتبر عاملا ثانياً) .
(جـ) مركبة عشوائية .

وتوضع استجابات وحدات التجريب كما يلي حيث $س_ر$ ، $ص_ر$ ترمزان إلى الاستجابات في القطاع $ر$ للمستوى الأول وللمستوى الثاني على الترتيب .

القطاع	المستوى الأول	المستوى الثانى
(١)	$س_١$	$ص_١$
(٢)	$س_٢$	$ص_٢$
(٣)	$س_٣$	$ص_٣$
....	....	....
(ن)	$س_ن$	$ص_ن$

بهذا التصور نكون بصدد حالة خاصة للتجارب ذوات العاملين غير المتفاعلين ، وبالتالى تسير عملية تحليل التباين كما في البند (٨ – ٦ – ١) السابق .

مثال (٨ – ٨) :

أراد باحث أن يعرف ما إذا كان عرض الجزء الأسفل من وجه البنات أكبر في سن السادسة منه في سن الخامسة ، فأختار عينة عشوائية من ١٥ بنتاً وقاس عرض الوجه وهن في الخامسة ثم أعاد القياس بعد عام ، وسجلت الأطوال بالسنتيمتر كما يلي :

جدول (٨ - ١٠)

الأفراد (القطاعات)	(١) ٥ سنوات	(٢) ٦ سنوات	$\text{م}_{\text{ر}}$
(١)	٧,٣٣	٧,٥٣	١٤,٨٦
(٢)	٧,٤٩	٧,٧٠	١٥,١٩
(٣)	٧,٢٧	٧,٤٦	١٤,٧٣
(٤)	٧,٩٣	٨,٢١	١٦,١٤
(٥)	٧,٥٦	٧,٨١	١٥,٣٧
(٦)	٧,٨١	٨,٠١	١٥,٨٢
(٧)	٧,٤٦	٧,٧٢	١٥,١٨
(٨)	٦,٩٤	٧,١٣	١٤,٠٧
(٩)	٧,٤٩	٧,٦٨	١٥,١٧
(١٠)	٧,٤٤	٧,٦٦	١٥,١٠
(١١)	٧,٩٥	٨,١١	١٦,٠٦
(١٢)	٧,٤٧	٧,٦٦	١٥,١٣
(١٣)	٧,٠٤	٧,٢٠	١٤,٢٤
(١٤)	٧,١٠	٧,٢٥	١٤,٣٥
(١٥)	٧,٦٤	٧,٧٩	١٥,٤٣
$\text{م}_{\text{ع}}$	١١١,٩٢	١١٤,٩٢	م=٢٢٦,٨٤، ن= ٣٠
$\bar{\text{س}}_{\text{ع}}$	٧,٤٦	٧,٦٦	$\bar{\text{س}}$ = ٧,٥٦
$\text{مج}_{\text{ر}}\ \text{س}^{٢}_{\text{رع}}$	٨٣٦,٣٣	٨٨١,٨٣٠٤	$\text{مج مج س}^{٢}_{\text{رع}}$=١٧١٨,١٦٠٤

لدينا عامل تجريب واحد هو العمر وهذا العامل له مستويان هما ٥ سنوات ، ٦ سنوات ونريد المقارنة بين متوسطى عرض الوجه في هذين المستويين . وتحسباً لوجود فروق بين أفراد العينة فإننا ندخل هؤلاء الأفراد كعامل ثان مع ملاحظة أن كل فرد يمثل

قطاعاً خضع لكل من مستويي العامل . وعلى أساس عدم وجود تفاعل بين العاملين نسير في تحليل التباين كما في البند (٨ – ٦ – ١) لنجد ما يلي :

م م (الكلي) = ١٧١٨,١٦٠٤ – $\frac{{٢٢٦,٨٤}^2}{٣٠}$ = ١٧١٨,١٦٠٤ – ١٧١٥,٢١٢٨

= ٢,٩٤٧٦ حيث ν = ٢٩

م م (بين العمرين) = $\frac{{١١١,٩٢}^2 + {١١٤,٩٢}^2}{١٥}$ – ١٧١٥,٢١٢٨

= ١٧١٥,٥١٢٨ – ١٧١٥,٢١٢٨

= ٠,٣٠٠٠ حيث $\nu_{٢}$ = ١

م م (بين الأفراد) = $\frac{{١٤,٨٦}^2 + {١٥,١٩}^2 + \cdots + {١٥,٤٣}^2}{٢}$ – ١٧١٥,٢١٢٨

= ٧١٧,٨٤٩٦ – ١٧١٥,٢١٢٨

= ٢,٦٣٦٨ حيث $\nu_{٣}$ = ١٤

م م (الخطأ) = ٢,٩٤٧٦ – (٠,٣٠٠٠ + ٢,٦٣٦٨)

= ٠,٠١٠٨ حيث $\nu_{٤}$ = ١٤

جدول (٨ – ١١)

مصدر التباين	م م	ν	تقدير التباين	ف
بين العمرين	٠,٣٠٠٠	١	٠,٣٠٠٠	٣٨٩,١١**
بين الأفراد	٢,٦٣٦٨	١٤	٠,١٨٨٣	٢٤٢,٠٢**
الخطأ	٠,٠١٠٨	١٤	٠,٠٠٠٧	
الكلي	٢,٩٤٧٦	٢٩		

الاستنتاج :

(١) نسبة التباين للأعمار ذات دلالة عالية لأن ف٠,٠١ [١ ، ١٤] أصغر من ٩ ، مما يجعلنا نستنتج أن عرض وجه البنات في سن السادسة أوسع منه في سن الخامسة .

(٢) كما أن الفروق بين أفراد العينة ذات دلالة عالية لأن ف٠,٠١ [١٤، ١٤] أصغر من ٤٠

ملاحظة (١١) :

إذا كنا لم ندخل في اعتبارنا الاختلافات بين أفراد العينة وأجرينا عملية تحليل التباين على أساس أن لدينا تجربة ذات عامل واحد كما في البند (٨ – ٤) نحصل على جدول التباين الآتي :

جدول (٨ – ١٢)

مصدر التباين	م م	ν	تقديرات التباين	ف
بين العمرين	٠,٣٠٠٠	١	٠,٣٠٠٠	٣,١٧
الخطـــأ	٢,٦٤٧٦	٢٨	٠,٠٩٤٦	
الكلي	٢,٩٤٧٦	٢٩		

وهنا نجد أن نسبة التباين للأعمار ليست ذات دلالة مما يشير إلى عدم وجود فرق بين عرض الوجه في العمرين الخامسة والسادسة . وهذه النتيجة خاطئة وتخالف ما توصلنا إليه من قبل والسبب في ذلك عدم تجانس أفراد العينة ووجود فروق

بينهن نشأ عنه اختلافات كبيرة كان يجب أن تستبعد ولكنها أضيفت إلى الاختلاف العشوائي فتضخم مجموع مربعات الخطأ وهذا بدوره دخل في مقام نسبة التباين فجعل هذه النسبة أصغر من أن تكون ذات دلالة ، وبالتالى اختفي الفرق بين مستويي عامل التجريب .

إن للمقارنات التزاوجية تطبيقات عديدة في مختلف ميادين البحث العلمى خاصة عند القياس أو الاختبار المتكرر لنفس المجموعة بعد فترة ما أو بعد حدث ما حيث يجرى القياس « قبل وبعد » هذه الفترة أو هذا الحدث . ومن أمثلة ذلك اختبار قوة عضلات مجموعة من الأفراد ثم تعريضهم لتمرينات رياضية عنيفة ثم اختبار قوة عضلاتهم مرة أخرى . كذلك قياس خاصة ما لمجموعة من الكائنات الحية أو الأفراد ثم قياس هذه الخاصة لنفس المجموعة بعد مرحلة ما كما في المثال (٨ – ٨) الأخير .

كذلك تنتج المقارنات التزاوجية عند تقسيم وحدة ما إلى نصفين يتلقي أحدهما أحد مستويي عامل ما ويتلقي النصف الآخر المستوى الثاني الذى يمكن أن يكون مستوى المراقبة . وكمثال لذلك اختبار قوة نوعين من المضادات الحيوية يحقن أحدهما في الذراع الأيمن والآخر في الذراع الأيسر لنفس الشخص ثم قياس قطر البقع الحمراء الناتجة في كل من الحالتين .

ومن التصميمات التي تؤدى إلى مقارنات تزاوجية أيضاً إعطاء معالجتين إلى شخصين يشتركان في خبرة واحدة سواء كانت خبرة وراثية أو بيئية ، كإعطاء دواء إلى مجموعة من التوائم أو الأشقاء يتلقي أحدهما الدواء ولا يتلقاه الآخر (مجموعة مراقبة) .

(٨ – ٧ – ١) اختبار ت للمقارنات التزاوجية :

ذكرنا في الملاحظة (٤) بالبند (٨ – ٤) أننا حين نتناول عاملا ذا مستويين (ك = ٢) يمكن أن نستخدم اختبار ت للمقارنة بين متوسطى هذين المستويين

وتكون النتيجة التي نحصل عليها من هذا الاختبار مطابقة تماماً للنتيجة التي نحصل عليها من تحليل التباين . غير أن الإحصاءة التي مرت بنا بالبند (٦ – ٦ – ٣) لا تصلح لهذا الغرض في حالة المقارنات التزاوجية لأن أحد شروط استخدام تلك الإحصاءة استقلال المجموعتين بينما نحن هنا بصدد مجموعتين مرتبطتين بل هما نفس المجموعة . أما الإحصاءة ت التي تصلح لذلك فتأخذ الصورة الآتية :

$$\text{ت} = \frac{\bar{\text{ف}} - (\mu_{1} - \mu_{2})}{\text{ع}_{\text{ف}} / \sqrt{\text{ن}}} \qquad \text{بدرجات حرية } \nu = \text{ن} - 1 \qquad (13)$$

$$\text{حيث } \text{ف}_{\text{ر}} = \text{س}_{\text{ر}} - \text{ص}_{\text{ر}}$$

$$\text{،}\ \bar{\text{ف}} = \frac{1}{\text{ن}} \sum \text{ف}_{\text{ر}} = \frac{1}{\text{ن}} \sum (\text{س}_{\text{ر}} - \text{ص}_{\text{ر}})$$

= متوسط الفروق بين استجابات المستويين

$$\text{،}\ \text{ع}^{2}_{\text{ف}} = \frac{1}{\text{ن} - 1} \sum (\text{ف}_{\text{ر}} - \bar{\text{ف}})^{2} = \text{تباين الفروق ف}_{\text{ر}}$$

$$= \frac{1}{\text{ن} - 1} \left[\sum \text{ف}^{2}_{\text{ر}} - \frac{(\sum \text{ف}_{\text{ر}})^{2}}{\text{ن}} \right] \qquad (14)$$

مع ملاحظة أن $\text{ع}_{\text{ف}} / \sqrt{\text{ن}}$ هو الخطأ المعيارى لمتوسط الفروق مقدراً من العينة. والاختبار الذى نجريه باستخدام الإحصاءة (١٣) يعطى نتيجة مطابقة تماماً لما تعطيه طريقة تحليل التباين إلا أنه لا يزودنا بمقياس لتباين القطاعات (الصفوف) . وفي المثال (٨ – ٨) الأخير يمكننا الحل كما يلى :

الجدول (٨ - ١٣)

الأفراد	ف = س − ص	ف²	الأفراد	ف = س − ص	ف²
١	٠,٢٠	٠,٠٤	٩	٠,١٩	٠,٠٣٦١
٢	٠,٢١	٠,٠٤٤١	١٠	٠,٢٢	٠,٠٤٨٤
٣	٠,١٩	٠,٠٣٦١	١١	٠,١٦	٠,٠٢٦٥
٤	٠,٢٨	٠,٠٧٨٤	١٢	٠,١٩	,٠٣٦١
٥	٠,٢٥	٠,٠٦٢٥	١٣	٠,١٦	٠,٠٢٥٦
٦	٠,٢٠	٠,٠٤	١٤	٠,١٥	٠,٠٢٢٥
٧	٠,٢٦	٠,٠٦٧٦	١٥	٠,١٥	٠,٠٢٢٥
٨	٠,١٩	٠,٠٣٦١			
				٣,٠٠	٠,٦٢١٦

$\bar{\text{ف}} = \frac{٣}{١٥} = ٠,٢٠$ من السنتيمتر

$\text{ع}^{٢}_{\text{ن}} = \frac{١}{١٤} (٠,٦٢١٦ - \frac{٩}{١٥}) = ٠,٠١٥٤٣$

$\text{ع}_{\text{ن}} = ٠,٠٣٩٢٨١٠$

$\text{ع}_{\text{ن}} / \sqrt{\text{ن}} = \frac{٠,٠٣٩٢٨١٠}{\sqrt{١٥}} = ٠,٠١٠١٤٢٣$

الفرض الصفرى $\text{ف}_{٠}$: $\mu_{١} = \mu_{٢}$

الفرض الآخر $\text{ف}_{١}$: $\mu_{٢} > \mu_{١}$

بحساب قيمة الإحصاءة (١٣) من بيانات العينة (وعلى أساس صحة الفرض الصفرى) نجد أن

$$ت_ى = \frac{٠,٢٠ - ٠}{٠,٠١٠١٤} = ١٩,٧٢٤^{**} \qquad \text{بدرجات حرية عددها ١٤}$$

وهذه القيمة ذات دلالة عالية مما يجعلنا نرفض الفرض الصفرى ونستنتج أن عرض وجه البنات أوسع فى سن السادسة منه فى سن الخامسة .

ملاحظات :

(١) يمكن أن نثبت رياضيا أن مربع قيمة المتغير ت عند درجة الحرية ن يساوى قيمة المتغير ف عند درجتى الحرية (١ ، ن) . ففى المثال الأخير نجد أن : $ت^٢_ى$ = (١٩,٧٢٤)^٢ = ٣٨٩,٠٣٦ . وهذا العدد يساوى العدد $ف_ى$ = ٣٨٩,١١ الذى ظهر بالجدول (٨ - ١١) فيما عدا الفرق الناشىء عن عمليات التقريب .

(٢) فى مثل هذه الحال يكون لدينا متغيران غير مستقلين س ، ص ، إلا أننا فى استخدام الإحصاءة (١٣) نعتبر أن لدينا متغيرا واحدا ف حيث ف = س - ص. ولما كان المتغيران س ، ص هما متغيران معتدلان متوسطاهما $\mu_١$ ، $\mu_٢$ وتبايناهما $\sigma^٢_١$ ، $\sigma^٢_٢$ على الترتيب ، فإن المتغير ف يكون متغيرا معتدلا متوسطه $\mu_ف = \mu_١ - \mu_٢$ وهذا المتوسط ينعدم إذا كان الفرض الصفرى صحيحا . أما التباين $\sigma^٢_ف$ فنقدره من العينة بالمقدار $ع^٢_ف$ المعرف فى (١٤) بصرف النظر عن تساوى أو عدم تساوى التباينين $\sigma^٢_١$ ، $\sigma^٢_٢$. أى أننا فى استخدامنا للإحصاءة (١٣) لا نكون بحاجة لفرض تساوى هذين التباينين .

(٣) إن حدى الثقة بدرجة ١ - α للفرق $\mu_١ - \mu_٢$ بين متوسطى مجتمعى المتغيرين غير المستقلين س ، ص هما (بالأسلوب المعتاد) :

$$\bar{ف} \pm \frac{ع_ف}{\sqrt{ن}} ت_{\alpha(ن-١)} \qquad (١٥)$$

ففى المثال الأخير نجد أن حدى الثقة بدرجة ٩٩٪ للفرق بين متوسط عرض وجه البنات فى سن السادسة ومتوسط عرضه فى سن الخامسة هما

٠,٢٠ ± ٠,٠١٠١٤ × ٢,٩٧٧ أى ٠,١٧٠ ، ٠,٢٣٠ من السنتيمترات

تمارين (٨ – ٤)

١ – أراد طبيب باحث أن يقرر ما إذا كان تعاطى قرص من مادة معينة يحدث تأثيرا جانبيا غير مرغوب فيه من حيث تخفيض ضغط الدم . وقد بدأت التجربة بقياس ضغط الدم لعينة عشوائية من ١٥ فردا ثم إعطاء الأقراص لأفراد هذه العينة ، وانتهت بقياس ضغط الدم مرة أخرى بعد فترة معينة . سجلت القياسات كما يلى حيث س تعبر عن الضغط قبل تعاطى القرص ، ص تعبر عن الضغط لنفس الشخص بعد تعاطى القرص .

الفـــرد :	(١)	(٢)	(٣)	(٤)	(٥)	(٦)	(٧)	(٨)	(٩)	(١٠)	(١١)	(١٢)	(١٣)	(١٤)	(١٥)
س :	٧٠	٨٠	٧٢	٧٦	٧٦	٧٦	٧٢	٧٨	٨٢	٦٤	٧٤	٩٢	٧٤	٦٨	٨٤
ص :	٦٨	٧٢	٦٢	٧٠	٥٧	٦٦	٦٨	٥٢	٦٤	٧٢	٧٤	٦٠	٧٤	٧٢	٧٤

هل هذه البيانات تؤدى إلى الحكم بأن للأقراص تأثيرا فى تخفيض ضغط الدم ؟ استخدم كلا من طريقة تحليل التباين واختبار ت وقارن بين النتيجتين .

٢ – كان أحد الأطباء يشك فى أن الميزان الذى يزن به المرضى فى عيادته يعطى قراءات أعلى من القراءات التى يجدها المرضى عند استخدامهم للموازين التى فى منازلهم . ولاختبار ذلك طلب الطبيب من عشرة مرضى تسجيل أوزانهم بملابسهم الكاملة قبل مغادرتهم منازلهم إلى عيادته ثم قام بوزنهم فور وصولهم فحصل على البيانات الآتية حيث س تعبر عن الوزن فى العيادة ، ص تعبر عن الوزن بالمنزل

س :	١٥٣	١٤٦	٢٠٧	١٧٣	١٢٧	١١٢	١٨١	١٣٢	١٣٩	٢١٣
ص :	١٥٠	١٤٧	٢٠٣	١٧١	١٢٩	١١٠	١٧٩	١٣٢	١٣٨	٢١٠

اختبر ما إذا كان الطبيب محقا فى شكه وأوجد فترة ثقة بدرجة ٩٥٪ للفرق بين نوعى الوزن .

(٨ – ٨) التجارب ذوات الثلاثة العوامل (في مربع لاتيني)
LATIN SQUARES

إن المشكلة الأساسية في تناول التجارب ذوات الثلاثة العوامل هى أنها تحتاج إلى عدد كبير من وحدات التجريب وينبغى البحث حينئذ عن تصميمات توفر في عدد هذه الوحدات ، وفكرة المربع اللاتيني تحقق هذا التوفير وتعطى مثالاً جيداً لأهمية اختيار التصميم الكفء أى الذى يعطى نتائج كثيرة بأقل قدر من الجهد التجريبي .

والمربع اللاتيني من الرتبة ل هو تنظيم لحروف ا ، ب ، جـ ، ... عددها ل على هيئة مصفوفة مربعة (ل × ل) يشترط فيها أن يقع كل حرف مرة واحدة بالضبط في كل صف ومرة واحدة بالضبط في كل عمود ، وبذلك يظهر كل حرف ل بالضبط من المرات . المصفوفات الآتية هى أمثلة لمربعات لاتينية ذوات الرتب الثالثة والرابعة والخامسة ، علماً بأنه يمكن كتابة أى منها بصور أخرى مختلفة .

ا	ب	حـ
ب	حـ	ا
حـ	ا	ب

ى	ا	حـ	ب
حـ	ب	ى	ا
ب	ى	ا	حـ
ا	حـ	ب	ى

ب	ى	هـ	ا	حـ
حـ	ا	ب	هـ	ى
ى	حـ	ا	ب	هـ
ا	هـ	ى	حـ	ب
هـ	ب	حـ	ى	ا

في أى مربع لاتيني لدينا أعمدة يمكن أن تمثل عناصرها مستويات عامل ما ،

وصفوف يمكن أن تمثل عناصرها مستويات عامل ثان ، ولدينا أيضاً حروف ا ، ب ، جـ ، . . . يمكن أن تمثل مستويات عامل ثالث . ولذلك تصلح المربعات اللاتينية لدراسة التجارب ذوات الثلاثة العوامل بطريقة تعد امتداداً للطريقة التي استخدمناها بالبند (٨ - ٦ - ١) . وبالاضافة إلى الشروط المعتادة ، ينبغى توفر الشرطين الآتيين لاستخدام المربعات اللاتينية :

١ - أن يتساوى عدد مستويات كل من العوامل الثلاثة ، وبذلك يكون عدد الأعمدة = عدد الصفوف = عدد الحروف = ل ، ويكون عدد وحدات التجريب هو ن = ل^2 .

٢ - ألا يكون هناك تفاعلات بين العوامل الثلاثة ، لأن هذا التصميم لا يقيس التفاعل نظرا لأن هناك مشاهدة واحدة فقط فى كل خلية .

(يحسن ألا تستخدم المربعات اللاتينية التي تقل رتبها عن خمسة لعدم وجود معلومات كافية عن مدى حساسية المربعات اللاتينية لانحراف ظروف التجربة عن الفروض الموضوعة) .

ولتحليل التباين من مربع لاتيني نوجد م^2 (الكلى) ، م^2 (بين الأعمدة) ، م^2 (بين الصفوف) كما في البند (٨ - ٦ - ١) وبالنسبة للعامل الثالث نوجد الاختلاف الناشيء عنه من الصيغة :

$$\text{م}^2\,(\text{الحروف}) = \frac{\text{مجـ}\ \text{م}_{\text{س}}^2}{\text{ل}} - \frac{\text{م}^2}{\text{ن}}$$

حيث $\text{م}_{\text{س}}$ هو مجموع القيم المرتبطة بالحرف س (س = ا ، ب ، حـ . . .) . أما م^2 (الخطأ) فيحسب بطرح مجموع الاختلافات الناشئة عن العوامل الثلاثة من الاختلاف الكلى .

مثال (٨ – ٩) :

لاختبار تأثير ٣ عوامل على المحصول وهى تأثير اختلاف الأسمدة (٥ أنواع) وتأثيرا اختلاف التربة في اتجاهين متعامدين ، قسمت قطعة أرض إلى ٢٥ حوضاً متساوية المساحة ومرتبة في ٥ صفوف موازية لأحد الاتجاهين ، ٥ أعمدة موازية للاتجاه المتعامد عليه ، ثم وزعت الأنواع الخمسة من الأسمدة عشوائيا على الأحواض بحسب خطة مربع لاتيني من الرتبة الخامسة ، ونتج المربع اللاتيني الآتي حيث تشير الحروف إلى الأسمدة وتشير الصفوف والأعمدة إلى الاتجاهين المتعامدين وتشير الأعداد إلى الكميات الناتجة من المحصول (مطروح ١٠ من كل منها) والمطلوب بحث تأثير كل من هذه العوامل الثلاثة .

					مر
−٢,٦ أ	−١,١ د	−٤,٢ هـ	٢,٠ ب	٤,٣ حـ	−١,٦
١,٨ ح	−٣,٥ ب	−١,٣ أ	−٢,٤ هـ	−٢,١ د	−٧,٥
٠,١ د	٧,٩ حـ	−١,٠ ب	−١,٥ أ	−٢,٩ هـ	٢,٦
−١,٢ هـ	٠,١ أ	٥,٧ حـ	١,١ د	−٢,٦ ب	٣,١
١,٨ ب	−١,٢ هـ	٤,٣ د	٨,٤ حـ	٠,١ أ	١٣,٤
مع −٠,١	٢,٢	٣,٥	٧,٦	−٣,٢	م=١٠

الحل :

بالنسبة للأسمدة (الحروف) نجد أن المجاميع مـ كما يلى :

$م_{س}$: أ = −٢,٥ ، ب= −٣,٣ ، جـ = ٢٨,١ ، د= ٢,٣ ، هـ = −١١,٩

م م (الكلى) = مج مج س$^{٢}_{ر ن}$ − $\frac{م^{٢}}{ن}$

= ٢٨٥,١٨ − $\frac{١٠٠}{٢٥}$ = ٢٨١,١٨ حيث $\nu_{١}$ = ٢٤

م م (بين الأعمدة) = مج $\frac{م^{٢}_{ع}}{ل}$ − $\frac{م^{٢}}{ن}$

= ١٧,٠٢ − ٤ = ١٣,٠٢ حيث $\nu_{٢}$ = ٤

م م (بين الصفوف) = مج $\frac{م^{٢}_{ص}}{ل}$ − $\frac{م^{٢}}{ن}$

= ٥٠,٩٥ − ٤ = ٤٦,٩٥ حيث $\nu_{٣}$ = ٤

م م (بين الأسمدة) = مج $\frac{م^{٢}_{س}}{ل}$ − $\frac{م^{٢}}{ن}$

= ١٩٤,٨٩ − ٤ = ١٩٠,٨٩ . حيث $\nu_{٤}$ = ٤

م م (الخطأ) = ٢٨١,١٨ − (١٣,٠٢+ ٤٦,٩٥+ ١٩٠,٨٩)

= ٣٠,٣٢ حيث $\nu_{ه}$ = ٢٤ − ١٢ = ١٢

$ف_{٠,٠٥[٤ ، ١٢]}$ = ٣,٣٦ ، $ف_{٠,٠١[٤ ، ١٢]}$ = ٥,٤١

الاستنتاج : انظر الجدول (٨ ــ ١٤) .

(١) الاختلاف الناشيء عن الأسمدة ذو دلالة عالية .

(٢) الاختلاف الناشيء عن الأعمدة ليس له دلالة عند المستوى ٠,٠٥

(٣) الاختلاف الناشيء عن الصفوف ذو دلالة عند المستوى ٠,٠٥ وليس ذى دلالة عند المستوى ٠,٠١

جدول (٨ – ١٤)

مصدر التباين	مجموع المربعات	درجات الحرية	تقدير التباين	ف
بين الأعمدة	١٣,٠٢	٤ = ل − ١	٣,٢٥٥	١,٣
بين الصفوف	٤٦,٩٥	٤ = ل − ١	١١,٧٣٧	٤,٦*
بين الأعمدة	١٩٠,٨٩	٤ = ل − ١	٤٧,٧٢٢	١٨,٩**
الخطأ	٣٠,٣٢	١٢ = (ل−١)(ل−٢)	٢,٥٢٧	
المجموع	٢٨١,١٨	٢٤ = ل٢ − ١		

تمارين (٨ – ٥)

(١) في دراسة في أبحاث السوق كان المطلوب اختبار تأثير ثلاثة عوامل على مبيعات نوع معين من الغذاء في قطر ما وهذه العوامل هي :

(١) طرق التعبئة – ٤ مستويات : أ ، ب ، جـ ، ى

(٢) المناطق المختلفة في القطر – ٤ مستويات .

(٣) طرق التشجيع على الشراء – ٤ مستويات : نسبة تخفيض ، يانصيب ، كوبونات ، هدايا .

وقد أجريت تجربة باستخدام تصميم مربع لاتيني من الرتبة الرابعة وسجلت المبيعات في أسبوع بعشرات الآلاف من الدولار كالآتي :

المناطق	تخفيض	يانصيب	كوبونات	هدايا	المجموع
(١)	٤٨ أ	٣٨ ب	٤٢ جـ	٥٣ د	١٨١
(٢)	٣٩ ب	٤٣ جـ	٥٠ د	٥٤ أ	١٨٦
(٣)	٤٢ جـ	٥٠ د	٤٧ أ	٤٤ ب	١٨٣
(٤)	٤٦ د	٤٨ أ	٤٦ ب	٥٢ جـ	١٩٢
المجموع	١٧٥	١٧٩	١٨٥	٢٠٣	٧٤٢

مج أ = ١٩٧ مج ب = ١٦٧ مج جـ = ١٧٩ مج د = ١٩٩

ابحث دلالة تأثير كل من العوامل الثلاثة .

(٢) في المثال (٨ – ٩) أثبت أن الخطأ المعيارى

$ع_ع \sqrt{\frac{١}{ن_١} + \frac{١}{ن_٢}} = ع_ع \sqrt{\frac{٢}{ل}}$ للفرق بين متوسطين هو ١,٠٠٥

ومن ثم بين أن الفرق الذى يقل عن ٢,٢ بين أى متوسطين لا يكون ذا دلالة عند المستوى ٠,٠٥ ومن ثم برهن أن :

(أ) متوسط المحصول الناتج من السماد جـ أكبر جوهرياً من ذلك الناتج من أى نوع آخر من السماد .

(ب) متوسط المحصول الناتج من السماد د أكبر جوهرياً من ذلك الناتج من السماد هـ .

(٨ – ٩) معالجة الانحرافات عن افتراضات التحليل :

إن سلامة ما نجريه من تحليل وما نخرج به من نتائج تتوقف على توفرالافتراضات المذكورة فى البند (٨ – ٤ – ١) . ولكن ماذا يكون موقفنا إذا لم تكن بعض هذه الافتراضات مستوفاة ؟ هذا ما سناقشه كما يلى :

(أ) افتراض الاعتدالية :

إن افتراض اعتدال المجتمعات التى نتناولها فى تحليل التباين هو افتراض رئيسى . ولكن نظرا لأننا فى هذا التحليل (بالنموذج ثابت التأثيرات) نبحث فى الفروق بين المتوسطات فإن هذاالافتراض يمكن أن نتحلل منه دون خطورة بشرط أن تكون العينات كبيرة كبرا كافيا (أى لا يقل حجم كل منها عن ٣٠) . وذلك لأنه حسب نظرية النهاية المركزية – انظر ملاحظة البند (٦ – ٣) – إذا كانت الاستنتاجات المتعلقة بالأوساط الحسابية صحيحة حين تكون المجتمعات معتدلة فإنها تظل صحيحة حين تكون المجتمعات غير معتدلة بشرط أن تكون العينات كبيرة كبرا كافيا . وكلما اشتد انحراف المجتمع عن الاعتدالية كلما وجب علينا زيادة حجوم العينات .

(ب) افتراض تساوى التباينات :

يتضمن الافتراض الثانى من افتراضات تحليل التباين أن تكون مجتمعات أقسام أو مستويات عامل التجريب متساوية التباين . يمكن التجاوز عن هذاالافتراض بشرط أن تكون العينات متساوية الحجم لأنه فى هذه الحال لا يكون التحليل حساسا للانحرافات الصغيرة عن افتراض تساوى التباينات . أما إذا لم تكن العينات متساوية الحجم فإن اختلاف تباينات مجتمعاتها يكون ذا عواقب وخيمة على صحة الاستنتاج . ولهذا يفضل سحب عينات متساوية الحجم كلما أمكن ذلك .

(جـ) افتراض استقلال الأخطاء :

يتطلب الافتراض الثالث أن تكون أخطاء التجريب مستقلة عن بعضها وعن مستويات المعالجة ، وهذا أمر بالغ الأهمية لسلامة استخدام اختبار ف فى تحليل التباين . وعدم توفر هذا الشرط يمكن أن يؤدى إلى خطأ جسيم فى الاستنتاج . ولذلك وجب اتخاذ القدر الكافى من الحيطة فى عملية التجريب بحيث نضمن استقلال المشاهدات داخل وبين المجموعات ، ويساعدنا على ذلك تطبيق مبدأ العشوائية فى كافة جوانب التجربة وقد سبق الإشارة إلى ذلك .

وفى كثير من الأحيان يمكن تصحيح بعض الانحرافات عن الفروض الموضوعة باستخدام أحد التحويلات المناسبة كالمذكورة فى البند (٤ — ٥) . على أنه إذا فشلنا فى توفير هذه الفروض فلا مفر من الالتجاء فى التحليل إلى أحد الطرق غير البارامترية التى سنتناولها فى الفصل الرابع عشر .

(٨ — ١٠) عودة إلى مقارنة المتوسطات :

نعلم أن تحليل التباين ما هو إلا الخطوة الأولى لدراسة نتائج التجربة وأن الخطوة الثانية (فى النموذج ثابت التأثيرات) هى مقارنة متوسطات أقسام المعالجة واختبار ما قد يكون بينها من فروق . ولقد قدمنا فى البند (٨ — ٥) طريقة لاختبار هذه المقارنات فى حالتى المقارنات القبلية والبعدية . ونقدم الآن أسلوبا أو مدخلا آخر يسفر عن نفس الصيغ والاختبارات السابق تقديمها ولكن فى صور مختلفة وبدلالة متغيرات أخرى ، وبالتالى يسفر عن نفس الاستنتاجات . على أن دراسة هذا المدخل تعمق مفهوم المقارنات وتضفى عليها معانى مفيدة . وفى هذا البند والبنود الثلاثة التالية نقدم بعض التعاريف الأساسية فى هذا المدخل ونبدأ بالتعريف الآتى :

تعريف (١) : المقارنة (أو المضادة)

COMPARISON (or CONTRAST)

(أ) المقارنة بين متوسطات مجتمعات :

لتكن μ_1 ، μ_2 ، . . . ، $\mu_ك$ متوسطات ك من المجتمعات . ولتكن $ا_1$ ، $ا_2$ ، . . . $ا_ك$ أعدادا ثابتة ليست جميعها أصفارا . إن أى تعبير خطى ψ فى هذه المتوسطات :

$$\psi = ا_1 \mu_1 + ا_2 \mu_2 + \ldots + ا_ك \mu_ك = \sum ا_ر \mu_ر \qquad (١٦)$$

يسمى مقارنة (خطية) بين متوسطات هذه المجتمعات إذا توفر الشرط الآتى :

$$ا_1 + ا_2 + \ldots + ا_ك = ٠ \text{ أى } \sum ا_ر = ٠$$

ويتضمن هذا الشرط أن تكون بعض المعاملات $ا_ر$ (وتسمى أوزانا) موجبة والبعض الآخر سالبة . والمعتاد اختيار هذه الأوزان فى أى مقارنة بحيث يكون مجموع الأوزان الموجبة مساويا للواحد وبالتالى يكون مجموع الأوزان السالبة مساويا للعدد −١ . وهذا الإجراء ممكن دائما وعيبه أنه يجعل الأوزان فى صور كسرية فى أغلب المقارنات ولذلك لا يفضله بعض الباحثين . ومن أمثلة المقارنات الخطية ما يلى :

$$\mu_1 - \mu_2 \text{ ، } \tfrac{1}{2}\mu_1 + \tfrac{1}{2}\mu_2 - \mu_3 \text{ ، } 2\mu_1 - \mu_2 - \mu_3 \text{ .}$$

(ب) المقارنة بين متوسطات عينات :

لتكن $\bar{س}_1$ ، $\bar{س}_2$ ، . . . ، $\bar{س}_ك$ متوسطات ك من العينات المأخوذة من

المجتمعات موضع الدراسة . ولتكن $\text{ا}_{١}$ ، $\text{ا}_{٢}$ ، . . . ، $\text{ا}_{ك}$ أعدادا ثابتة ليست جميعها أصفارا . إن أى تعبير خطى $\hat{\psi}$ فى هذه المتوسطات :

$$\hat{\psi} = \text{ا}_{١}\,\bar{\text{س}}_{١} + \text{ا}_{٢}\,\bar{\text{س}}_{٢} + \ldots + \text{ا}_{ك}\,\bar{\text{س}}_{ك} \qquad (١٧)$$

$$\text{حيث مج ا}_{ى} = ٠ \qquad (١٨)$$

يسمى مقارنة بين متوسطات هذه العينات .

ويمكن إثبات أن توفر الشرط (١٨) يجعل قيمة أى مقارنة بين متوسطات العينات مستقلة عن قيمة المتوسط العام $\bar{\bar{\text{س}}}$ لهذه العينات . وهذا أمر هام لأننا فى مقارنة متوسطات المجتمعات ينبغى ألا يكون لمتوسطات العينات أى علاقة بقيمة $\bar{\bar{\text{س}}}$ التى نستخدمها لتقدير المتوسط العام μ للمجتمع .

(حـ) المقارنة بين مجاميع عينات

بالمثل ، لتكن $\text{م}_{١}$ ، $\text{م}_{٢}$ ، . . . ، $\text{م}_{ك}$ مجاميع ك من العينات ولتكن $\text{ا}_{١}$ ، $\text{ا}_{٢}$ ، ... ، $\text{ا}_{ك}$ أعدادا ثابتة ليس جميعها أصفارا . إن أى تعبير خطى $\hat{\psi}$ فى هذه المجاميع :

$$\hat{\psi} = \text{ا}_{١}\,\text{م}_{١} + \text{ا}_{٢}\,\text{م}_{٢} + \ldots + \text{ا}_{ك}\,\text{م}_{ك} \qquad (١٩)$$

$$\text{حيث مج ا}_{ى} = ٠$$

يسمى مقارنة بين مجاميع هذه العينات .

وفى تحليل التباين ، المعتاد استخدام المقارنات بين المجاميع وليس بين المتوسطات لأن ذلك يخفف من بعض عمليات القسمة والتقريب . غير أننا سنتناول هنا المقارنات بين المتوسطات لأننا أساسا ندرس هذه المتوسطات وبالتالى فإن تناولها يكون أكثر قدرة

على ما نريد توضيحه من مفاهيم . وعلى أية حال فإن ما سنقدمه من صيغ تخص المقارنات بين المتوسطات يمكن تحويلها إلى صيغ تخص المقارنات بين المجاميع بمجرد ضرب كل معامل لر فى الحجم نر للعينة الخاصة به .

(يلاحظ أن رمز المقارنة ψ أو $\hat{\psi}$ هو عبارة عن عدد واحد لأنه يتركب من مجموع أعداد) .

إن أى تساؤل عن بعض أو كل متوسطات أقسام المعالجة يمكن وضعه فى صورة مقارنة بين هذه المتوسطات ولا يستلزم الأمر إلا وضع وزن مناسب لكل متوسط بحيث يتحقق الشرط (١٨) ، مع ملاحظة إعطاء الوزن صفر للمتوسطات التى لا تدخل فى المقارنة .

مثال (٨ – ١٠) :

فى المثال (٨ – ١) كان لدينا خمسة أقسام حجم كل منها ١٠ ومتوسطاتها كالآتى :

(١) جلوكوز	(٢) فركتوز	(٣) جلوكوز + فركتوز	(٤) سكروز	(٥) مراقبة
٥٩,٣	٥٨,٢	٥٨	٦٤,١	٧٠,١

وفى الأمثلة (٨ – ٢) و(٨ – ٣) و(٨ – ٤) التابعة لهذا المثال كان المطلوب الإجابة عن التساؤلات القبلية الآتية :

(١) هل متوسط مجموعة السكريات الأربعة مجتمعة يختلف عن متوسط مجموعة المراقبة ؟

(٢) هل متوسط مجموعة السكريات النقية (١) و(٢) و(٤) مجتمعة يختلف عن متوسط السكر الخليط ؟

(٣) هل متوسطات مجموعات السكريات (١) و(٢) و(٤) واحدة ؟

إن هذه التساؤلات يمكن صياغتها على هيئة مقارنات بين المتوسطات كالآتى ، مع ملاحظة أنه يمكن إعطاء أوزانا أخرى تتناسب مع الأوزان المأخوذة هنا ومع ملاحظة ضرورة توفر الشرط (١٨) :

$$\psi_1 = \frac{1}{4}\mu_1 + \frac{1}{4}\mu_2 + \frac{1}{4}\mu_3 + \frac{1}{4}\mu_4 - \mu_5$$

$$\psi_2 = \frac{1}{3}\mu_1 + \frac{1}{3}\mu_2 - \mu_3 + \frac{1}{3}\mu_4 - \text{صفر} \times \mu_5$$

أما التساؤل الثالث فيتضمن مقارنتين يمكن كتابتهما بصور مختلفة منها :

$$\psi_3 = \mu_1 - \mu_2 \text{ ،}$$
$$\psi_4 = \frac{1}{2}(\mu_1 + \mu_2) - \mu_4$$

وعادة ما تكتب المقارنات المناظرة بين متوسطات العينات فى جدول كالآتى :

جدول (٨ - ١٥)

المقارنات المتعلقة بالأمثلة (٨ - ٢) و(٨ - ٣) و(٨ - ٤)

المقارنة \ المتوسط	(١) ٥٩,٣	(٢) ٥٨,٢	(٣) ٥٨	(٤) ٦٤,١	(٥) ٧٠,١
$\hat{\psi}_1$: سكريات ضد مراقبة	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	١-
$\hat{\psi}_2$: سكريات نقية ضد سكر خليط	$\frac{1}{3}$	$\frac{1}{3}$	١-	$\frac{1}{3}$	٠
$\hat{\psi}_3$: جلوكوز ضد فركتوز	١	١-	٠	٠	٠
$\hat{\psi}_4$: (جلوكوز+فركتوز) ضد سكروز	$\frac{1}{2}$	$\frac{1}{2}$	٠	١-	٠

من هذا الجدول يسهل إيجاد قيم المقارنات الأربع كالآتى :

$$\hat{\psi}_{١} = \frac{١}{٤}\,(٥٩,٣ + ٥٨,٢ + ٥٨ + ٦٤,١) - ٧٠,١ = -١٠,٢$$

$$\hat{\psi}_{٢} = \frac{١}{٣}\,(٥٩,٣ + ٥٨,٢ + ٦٤,١) - ٥٨ + \text{صفر} = ٢,٥٣٣٣$$

$$\hat{\psi}_{٣} = ٥٩,٣ - ٥٨,٢ + ٠ + ٠ + ٠ = ١,١$$

$$\hat{\psi}_{٤} = \frac{١}{٢}(٥٩,٣ + ٥٨,٢) + ٠ - ٦٤,١ + ٠ = - ٥,٣٥$$

(٨ -١١) المقارنات القبلية :

كما سبق القول ، يفضل اختيار المقارنات القبلية بحيث تكون مستقلة إحصائيا عن بعضها ، وذلك لكى تكون المعلومات الناتجة من أى مقارنة ذات قيمة تخصها وحدها ولا تتداخل مع المعلومات الناتجة من أى مقارنة أخرى . ولذلك يهمنا أن نعرف قاعدة نستدل بها على استقلال أو عدم استقلال المقارنات .

(٨ - ١١ - ١) معيار استقلال مقارنتين :

يمكن البرهنة رياضيا على النظرية الآتية :

لتكن $\text{س}_{١}$ ، $\text{س}_{٢}$ ، . . . ، $\text{س}_{\text{ك}}$ هى ك من المتغيرات العشوائية المستقلة التى لها توزيعات معتدلة وتباين مشترك . ولتكن $\psi_{١}$ و$\psi_{٢}$ هما المقارنتان :

$$\psi_{١} = \text{ا}_{١}\,\text{س}_{١} + \text{ا}_{٢}\,\text{س}_{٢} + \ldots + \text{ا}_{\text{ك}}\,\text{س}_{\text{ك}}$$

$$\psi_{٢} = \text{ب}_{١}\,\text{س}_{١} + \text{ب}_{٢}\,\text{س}_{٢} + \ldots + \text{ب}_{\text{ك}}\,\text{س}_{\text{ك}}$$

يكون المتغيران العشوائيان $\psi_{١}$ و$\psi_{٢}$ مستقلين إحصائيا إذا توفر الشرط الآتى :

$$\text{ا}_{١}\,\text{ب}_{١} + \text{ا}_{٢}\,\text{ب}_{٢} + \ldots + \text{ا}_{\text{ك}}\,\text{ب}_{\text{ك}} = ٠$$

أى

$$\sum \text{ا}_{\text{ر}}\,\text{ب}_{\text{ر}} = ٠ \qquad (٢٠)$$

(لاحظ أن هذا الشرط يعتمد فقط على الأوزان ولا يعتمد على المتغيرات) .

إن هذه النظرية تمنحنا قاعدة أو معيارا للكشف عن استقلال المقارنات . فإذا كان لدينا ك من المجتمعات المعتدلة التى تشترك فى التباين وسحبنا منها ك من العينات المستقلة التى لها **نفس الحجم** فإنه تطبيقا لهذه النظرية تكون أى مقارنتين $\hat{\psi}_1$ و$\hat{\psi}_2$ بين متوسطات العينات مستقلتين إذا توفر الشرط (٢٠) .

ويقال لمقارنتين ينطبق عليهما معيار الاستقلال إنهما متعامدتان **orthogonal** . وتعامد مقارنتين هنا يعنى استقلالهما إحصائيا ، ولذلك تستخدم كلمتا « التعامد » و« الاستقلال » كمترادفتين ما دمنا نتعامل مع متغيرات مستقلة وتوزيعات معتدلة تشترك فى التباين .

وحين تكون العينات مختلفة الأحجام ، $\text{ن}_\text{ر}$ ترمز إلى حجم العينة ر فإن معيار الاستقلال يتخذ الصيغة الآتية :

$$\frac{\text{أ}_1 \text{ب}_1}{\text{ن}_1} + \frac{\text{أ}_2 \text{ب}_2}{\text{ن}_2} + \ldots + \frac{\text{أ}_\text{ك} \text{ب}_\text{ك}}{\text{ن}_\text{ك}} = 0$$

أى

$$\sum \frac{\text{أ}_\text{ر} \text{ب}_\text{ر}}{\text{ن}_\text{ر}} = 0 \qquad (٢١)$$

مثال (٨ – ١١) :

اختبر استقلال المقارنات المدونة بالجدول (٨ – ١٥) ، على فرض توفر شروط استقلال المتوسطات واعتدال المجتمعات وتساوى تبايناتها .

الحل :

بالجدول أربع مقارنات وإذن هناك $\text{ق}^4_2 = 6$ أزواج من المقارنات يراد اختبار استقلالها . ومع ملاحظة تساوى أحجام العينات نجد من الصيغة (٢٠) ما يلى :

للمقارنتين $\hat{\psi}_1$ و$\hat{\psi}_2$: $\frac{1}{4}\times\frac{1}{3} + \frac{1}{4}\times\frac{1}{3} + \frac{1}{4}\times(-1) + \frac{1}{4}\times\frac{1}{3} + (-1)\times 0 = 0$

وإذن هاتان المقارنتان مستقلتان .

للمقارنتين $\hat{\psi}_1$ و $\hat{\psi}_3$: $\frac{١}{٤} \times ١ + \frac{١}{٤} \times (-١) + ٠ = ٠$

وإذن هاتان المقارنتان مستقلتان .

بالمثل نجد أن الأزواج الأربعة الباقية مستقلة ، وبذلك تكون المقارنات الستة مستقلة مثنى مثنى . تحقق من ذلك .

(٨ – ١١ – ٢) اختبار المقارنات القبلية :

يعتمد اختبار المقارنات القبلية على الحقيقتين الآتيتين اللتين يمكن إثباتهما رياضيا .

(أولا) المقارنة $\hat{\psi}$ بين متوسطات العينات هي تقدير غير متحيز للمقارنة ψ بين متوسطات المجتمعات التي أخذت منها العينات والتي تحمل نفس الأوزان .

(ثانيا) على فرض استقلال المتوسطات ، وإذا كان σ^2 هو التباين المشترك للمجتمعات فإن الخطأ المعياري للمقارنة $\hat{\psi}$ بين المتوسطات هو

$$\sigma \sqrt{\text{مجـ} \frac{أ_ر^2}{ن_ر}} \qquad (٢٢)$$

حيث $ن_ر$ حجم العينة ر .

ونظرا لأن التباين σ^2 يكون عادة غير معروف فإننا نقدره من البيانات المشاهدة بنفس طريقة التقدير في تحليل التباين أي بواسطة $\hat{ع}^2_خ$ وهو تباين خطأ التجريب . وبذلك يكون تقدير الخطأ المعياري للمقارنة $\hat{\psi}$ بين المتوسطات هو $ع_{\psi}$ حيث :

$$ع_{\psi} = ع_خ \sqrt{\text{مجـ} \frac{أ_ر^2}{ن_ر}} \qquad (٢٣)$$

من هاتين الحقيقتين وحين تتوفر شروط اعتدالية المجتمعات وتساوى تبايناتها واستقلال المتوسطات – وهذا ما نفترضه عادة فى تحليل التباين – فإنه حسب البند (٦ – ٦) يكون للإحصاءة

$$ت = \frac{\hat{\psi} - \psi}{ع_{\psi}} \qquad (٢٤)$$

توزيع ت بدرجات حرية عددها هو عدد درجات حرية $ع^{٢}_{ع}$ والذى نرمز له بالرمز $\nu_{ع}$.

كما أنه من البند (٦ – ٦ – ٢) تكون الفترة

$$(\hat{\psi} - ع_{\psi} \cdot ت_{\alpha[\nu ع]} ، \hat{\psi} + ع_{\psi} \cdot ت_{\alpha[\nu ع]}) \qquad (٢٥)$$

هى فترة ثقة بدرجة (١ – α) للمقارنة ψ .

كل من اختبار ت بالصورة (٢٤) والفترة (٢٥) يصلح لاختبار أى فرض عن قيمة المقارنة ψ . وبالرغم من أن اختبار ت هو أصلا ، كما نعلم ، اختبار للفرق بين متوسطى عينتين مستقلتين أى للمقارنات التى على الصورة $\hat{\psi} = \bar{س}_{ر} - \bar{س}_{و}$ إلا أنه يصلح هنا أيضا لاختبار أى مقارنة مهما كان عدد المتوسطات الداخلة فيها . وهذا استثناء فى استخدام اختبار ت السابق دراسته . وبالنسبة للفترة (٢٥) ، إذا افترضنا قيمة معينة ا لمقارنة ψ ولم تقع هذه القيمة فى هذه الفترة فإننا ، كالمعتاد ، نرفض الفرض ψ = ا عند مستوى الدلالة α (اختبار ذو جانبين) .

وفى معظم الحالات يكون المطلوب اختبار الفرض الصفرى $\psi = ٠$ ضد الفرض $\psi \neq ٠$. وفى هذه الحالة تأخذ الإحصاءة (٢٤) الصيغة الآتية :

$$\text{ت} = \frac{\hat{\psi}}{\text{ع}_{\psi}} \qquad \text{بدرجة حرية } \nu_{\text{خ}} \qquad (٢٦)$$

$$\text{أو الصيغة المكافئة ف} = \frac{\hat{\psi}^{٢}}{\text{ع}^{٢}_{\psi}} \qquad \text{بدرجتى حرية ١ ، } \nu_{\text{خ}} \qquad (٢٧)$$

وفى هذه الحالة أيضا نرفض الفرض الصفرى $\psi = ٠$ عند مستوى الدلالة α إذا كانت الفترة (٢٥) لا تحتوى الصفر .

مثال (٨ – ١٢) :

أجب عن التساؤل الأول المطروح بالمثال (٨ – ١٠) مستخدما مستوى الدلالة ٠,٠١ مع تذكر أننا وجدنا فى تحليل التباين أن تباين خطأ التجريب هو $\text{ع}^{٢}_{\text{خ}} = ٥,٤٦$ بدرجات حرية $\nu_{\text{خ}} = ٤٥$ وأن كلا من المتوسطات بنى على ١٠ مشاهدات

الحل :

التساؤل المطلوب يشير إلى المقارنة $\psi_{١}$ بين متوسط أقسام السكريات مجتمعة ومتوسط قسم المراقبة .

الفرض الصفرى : $\psi_{١} = ٠$ والفرض الآخر $\psi_{١} \neq ٠$

وجدنا أن $\hat{\psi}_{١} = -١٠,٢$

من (٢٣) : $\text{ع}^{٢}_{\psi} = \text{ع}^{٢}_{\text{خ}} \sum \frac{\text{أ}^{٢}}{\text{ن}}$ حيث $\nu = ٤٥$

$$= ٥,٤٦ \times \frac{١}{١٠}\left[(-\tfrac{١}{٤})^{٢} + (-\tfrac{١}{٤})^{٢} + (-\tfrac{١}{٤})^{٢} + (-١)^{٢}\right]$$

$$= ٠,٦٨٢٥$$

من (٢٧) : $ف_ي = \frac{(-٢,١٠)^٢}{٠,٦٨٢٥} = ٥٢,٤٤٨١^{**}$

وبما أن $ف_{٠,٠١[١ ، ٤٥]}$ = ٧,٣١ نرفض الفرض الصفرى عند المستوى ٠,٠١ . يلاحظ أن قيمة $ف_ي$ الناتجة هي نفس قيمة $ف_ي$ التى سبق أن توصلنا إليها فى المثال (٨ – ٢) .

ويمكن بطبيعة الحال استخدام اختبار ت بالصيغة (٢٦) حيث نجد أن القيمة التى تنتج تساوى الجذر التربيعى لقيمة $ف_ي$ التى حصلنا عليها . كما يمكن استخدام فترة الثقة (٢٥) كالآتى :

$$\text{الحد الأدنى للفترة} = -١٠,٢ - \sqrt{٠,٦٨٢٥} \times ٢,٦٣ = -١٢,٣٧$$

$$\text{الحد الأعلى للفترة} = -١٠,٢ + \sqrt{٠,٦٨٢٥} \times ٢,٦٣ = -٨,٠٣$$

إذن الفترة (-١٢,٣٧ ، -٨,٠٣) هى فترة ثقة بدرجة ٩٩٪ للمقارنة $\psi_١$. وبما أنها لا تحتوى الصفر نرفض الفرض $\psi_١$ = ٠ عند المستوى ٠,٠١ .

إن الصيغ (٢٤) و(٢٥) و(٢٦) و(٢٧) لاختبار المقارنات القبلية التى وضعت أصلا للتجارب ذوات العامل الواحد تصلح كذلك للتجارب ذوات العاملين أو أكثر طالما كان النموذج المستخدم هو النموذج ثابت التأثيرات .

مثال (٨ – ١٣) :

أجب عن التساؤلات القبلية الثلاثة المطروحة بالمثال (٨ – ٧) علما بأن $ع^٢_خ$ = ٣,١٣ بدرجات حرية ٤٥ وأن كلا من المتوسطات بنى على ٨ مشاهدات .

الحل :

نضع هذه التساؤلات على هيئة مقارنات كما بالجدول (٨ – ١٦) الآتى :

جدول (٨-١٦)

المقارنات المتعلقة بالمثال (٨-٧)

المقارنة \ المتوسط	(١) ١٢,٦٢٥	(٢) ٥,٥	(٣) ٣,٨٧٥	(٤) ٢,٢٥
$\hat{\psi}_1$: درجة الحرارة ٨ ضد درجة الحرارة ٢٨	١	٠	٠	١-
$\hat{\psi}_2$: درجة الحرارة ١٤ ضد درجة الحرارة ٢١	٠	١	١-	٠
$\hat{\psi}_3$ الدرجتان (٨،٢٨) ضد الدرجتان (١٤،٢١)	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$

$\hat{\psi}_1$ = ١ × ١٢,٦٢٥ − ١ × ٢,٢٥ = ١٠,٣٧٥

$ع^2_{\psi_1}$ = ٣,١٣ × $\frac{1}{8}$ (١+١) = ٠,٧٨٢٥

$ف_ى$ = (١٠,٣٧٥)^2 / ٠,٧٨٢٥ = ١٣٧,٥٥٩**

بما أن $ف_{[٤٥،١]٠,٠١}$ = ٧,٣١ فإن $ف_ى$ تكون ذات دلالة عالية
ونرفض الفرض الصفرى أن ψ_1 = ٠
بالمثل نجد أن

$\hat{\psi}_2$ = ١,٦٢٥ و$ع^2_{\psi_2}$ = ٠,٧٨٢٥ و$ف_ى$ = ٣,٣٧٥

وإذن نقبل الفرض أن ψ_2 = ٠
كذلك

$\hat{\psi}_3$ = ٢,٧٥ و$ع^2_{\psi_3}$ = ٠,٣٩١٢٥ و$ف_ى$ = ١٩,٣٢٩**

وإذن نرفض الفرض أن ψ_3 = ٠
لاحظ أن هذه هى نفس النتائج التى توصلنا إليها فى المثال (٨ – ٧) .

(٨ - ١٢) تجزىء مجموع المربعات بين أقسام المعالجة :

إن تحليل التباين بالنموذج ثابت التأثيرات يكافىء فصل البيانات إلى مجموعة من المقارنات المستقلة ، إذ أن كل درجة من درجات الحرية تصاحب معالجة ما يناظرها مقارنة ما بين المتوسطات . ولتوضيح ذلك نبدأ بالتعريف الآتى :

تعريف (٢) : مجموع مربعات مقارنة :

إذا كانت $\hat{\psi} = a_1 \bar{y}_1 + a_2 \bar{y}_2 + \ldots + a_k \bar{y}_k$ مقارنة بين متوسطات العينات فإن مجموع مربعات هذه المقارنة يعرف كالآتى :

$$SS(\hat{\psi}) = (\hat{\psi})^2 / \sum \frac{a^2}{n} \qquad (٢٨)$$

ويمكن إثبات أن لهذا المجموع درجة واحدة من درجات الحرية .

فمثلا ، فى المثال (٨ - ٢) الذى يتناول المقارنة $\hat{\psi}_1$ بين أقسام السكريات مجتمعة ضد قسم المراقبة نجد ما يلى :

$$\hat{\psi}_1 = -١٠,٢$$

$$\sum \frac{a^2}{n} = \frac{1}{10}[(\frac{1}{4})^2 + (\frac{1}{4})^2 + (\frac{1}{4})^2 + (\frac{1}{4})^2 + 1^2] = ٠,١٢٥$$

$\therefore SS(\hat{\psi}_1) = (-١٠,٢)^2 / ٠,١٢٥ = ٨٣٢,٣٢$ بدرجة حرية واحدة .

وهذا العدد بالضبط هو الذى وجدناه فى المثال (٨ - ٢) .

كذلك ، فى المثال (٨ - ٣) الذى يتناول المقارنة $\hat{\psi}_2$ بين السكريات النقية والسكر الخليط نجد ما يلى :

$$\hat{\psi}_2 = ٢,٥٣٣٣ ، \sum \frac{a^2}{n} = ٠,١٣٣٣٣ ، SS(\hat{\psi}_2) = ٤٨,١٣$$

بدرجة حرية واحدة . وهذا العدد هو بالضبط العدد الذى وجدناه فى المثال (٨ – ٣) .

ويترتب على التعريف (٢) أن

$$\text{تباين}(\hat{\psi}) = \text{م م}(\hat{\psi}) \div ١ = \text{م م}(\hat{\psi})$$

وحين تتوفر الشروط المعتادة لتحليل التباين يكون توزيع نسبة التباين لأى مقارنة $\hat{\psi}$ وهى

$$\text{ف} = \frac{\text{م م}(\hat{\psi})}{\text{ع}^{٢}_{\text{ع}}} \qquad (٢٩)$$

مطابقا لتوزيع ف بدرجتى حرية ١ ، $\nu_{\text{ع}}$ ويمكن باستخدام (٢٨) ، (٢٣) إثبات أن هذه النسبة هى بذاتها نسبة التباين (٢٧) وهى $\text{ف} = \frac{\hat{\psi}^{٢}}{\text{ع}^{٢}_{\psi}}$.

نبحث الآن فى المقارنات المستقلة ، وفى هذا البحث تلزمنا القاعدتين الآتيتين اللتين يمكن إثباتهما رياضيا .

قاعدة (١) :

إذا كانت $\hat{\psi}_{١}$ و $\hat{\psi}_{٢}$ مقارنتين مستقلتين على نفس البيانات فإن :

$$\text{م م}(\hat{\psi}_{١} \text{ و } \hat{\psi}_{٢}) = \text{م م}(\hat{\psi}_{١}) + \text{م م}(\hat{\psi}_{٢}) \qquad (٣٠)$$

أى أن مجموع مربعات مقارنتين مستقلتين يساوى مجموع مربعات إحداهما مضافا إلى مجموع مربعات الأخرى . كما أن مجموع المربعات الناتج من ضم المقارنتين يكون له درجتان من درجات الحرية .

وتمتد هذه القاعدة لأى عدد منتهى من المقارنات المستقلة ولذلك نقول إن المقارنات المستقلة تجميعية additive

فمثلا ، فى المثال (٨ – ٤) الذى يتساءل عما إذا كان هناك فروق بين السكريات النقية (١) و(٢) و(٤) ، رأينا فى المثال (٨ – ١٠) أن هذا التساؤل يتضمن المقارنتين $\hat{\psi}_3$ و$\hat{\psi}_4$. ونظرا لاستقلال هاتين المقارنتين يجب حسب القاعدة (١) أن يكون المجموع م م ($\hat{\psi}_3$) + م م ($\hat{\psi}_4$) مساويا لمجموع المربعات الذى حصلنا عليه فى المثال (٨ – ٤) . بالحساب نجد ما يلى :

$\hat{\psi}_3$ = ١,١ ، مج $\frac{c^2}{n}$ = $\frac{1}{10}$ (١ + ١) = ٠,٢٠

م م ($\hat{\psi}_3$) = $(١,١)^2$ ÷ ٠,٢٠ = ٦,٠٥ بدرجة حرية واحدة

كذلك ، $\hat{\psi}_4$ = –٥,٣٥ ، مج $\frac{c^2}{n}$ = ٠,١٥ ، م م ($\hat{\psi}_4$) = ١٩٠,٨٢

إذن م م ($\hat{\psi}_3$) + م م ($\hat{\psi}_4$) = ٦,٠٥ + ١٩٠,٨٢

= ١٩٦,٨٧ بدرجات حرية عددها ٢

وهذا بالضبط ما وجدناه بالمثال (٨ – ٤) وبذلك يتحقق أن

م م ($\hat{\psi}_3$ و $\hat{\psi}_4$) = م م ($\hat{\psi}_3$) + م م ($\hat{\psi}_4$)

قاعدة (٢) :

إذا كان بتجربة ما ك من أقسام المعالجة وكان هناك ك – ١ من المقارنات المستقلة $\hat{\psi}_1$ ، $\hat{\psi}_2$ ، . . . ، $\hat{\psi}_{k-1}$ بين متوسطات العينات فإن :

م م (جميع $\hat{\psi}_i$) أى [م م ($\hat{\psi}_1$) + م م ($\hat{\psi}_2$) + . . . + م م ($\hat{\psi}_{k-1}$)]

= م م (بين أقسام المعالجة) (٣١)

وتتضمن هذه القاعدة أن مجموع مربعات أى مقارنة هو جزء من مجموع المربعات بين أقسام المعالجة (بدرجة حرية واحدة) ، كما تتضمن أنه لا يمكن أن يزيد عدد المقارنات المستقلة عن ك − ١ مقارنة . وهناك حرية كبيرة فى اختيار مجموعة من ك − ١ من المقارنات المستقلة ويتم هذا الاختيار بحسب التساؤلات التى تجرى التجربة للإجابة عنها ، فإذا بدأنا بمقارنة معينة يمكن دائما تكوين الـ ك − ٢ من المقارنات المستقلة الأخرى .

فمثلا ، اعتبر المثال (٨ − ١) والأمثلة الثلاثة التابعة له .

عدد أقسام المعالجة = ك = ٥

من المثال (٨ − ١) وجدنا أن م م (بين الأقسام) = ١٠٧٧,٣٢ بدرجات حرية ٤ .

المقارنات $\hat{\psi}_1$ و $\hat{\psi}_2$ و $\hat{\psi}_3$ و $\hat{\psi}_4$ المدونة بالجدول (٨ − ١٤) هى مقارنات مستقلة عددها ٤ (= ك − ١) .

م م ($\hat{\psi}_1$ و $\hat{\psi}_2$ و $\hat{\psi}_3$ و $\hat{\psi}_4$) = ٨٣٢,٣٢ + ٤٨,١٣ + ٦,٠٥ + ١٩٠,٨٢

= ١٠٧٧,٣٢ بدرجات حرية ٤

= م م (بين الأقسام)

وبذلك تتحقق القاعدة (٢) .

ملاحظة :

إن كل ما ذكر فى هذا البند عن المقارنات المستقلة بين المتوسطات ينطبق على المقارنات المستقلة بين مجاميع العينات ، مع ملاحظة أن مجموع مربعات مقارنة بين المجاميع هى $(\hat{\psi})^2 / \sum \text{ن}_\text{ر} \, \text{ا}^2_\text{ر}$ حيث $\hat{\psi}$ مقارنة بين المجاميع ، $\text{ن}_\text{ر}$ حجم العينة ر .

مثال (٨ – ١٤) :

وجدت المشاهدات الآتية فى تجربة ما :

جدول (٨ – ١٧)

	الأقسام (١)	(٢)	(٣)	(٤)	
	٤	٢	٣	٣	
	٦	٦	٥	٥	
	٥	٣	٧	٤	
		٥	٧	٥	
			٨	٣	
$ن_ر$	٣	٤	٥	٥	ن = ١٧
$م_ر$	١٥	١٦	٣٠	٢٠	م = ٨١
$\bar{س}_ر$	٥	٤	٦	٤	$\bar{س}$ = ٤,٧٦

(أولا) أوجد ٢ ٣ (بين الأقسام) ،

(ثانيا) حدد أوزانا لكل من المقارنات القبلية الثلاث الآتية بحيث تكون مستقلة ،

(أ) المجموعة (١) ضد المجموعة (٢) .

(ب) المجموعتان (١) و(٢) معا ضد المجموعة (٣)

(حـ) المجموعات (١) و(٢) و(٣) ضد المجموعة (٤)

(ثالثا) أوجد قيمة كل مقارنة وقيمة مجموع مربعاتها ثم اختبر دلالتها عند المستوى ٠,٠٥ علما بأن $ع^{2}_{ع}$ = ٢,٤٦ بدرجات حرية ١٣ . اكتب جدول التباين بالتفصيل .

الحل :

(أولا) $م^2$ (بين الأقسام) $= \frac{15^2}{3} + \frac{16^2}{4} + \frac{30^2}{5} + \frac{20^2}{5} - \frac{81^2}{17}$

= ١٣,٠٥٨ بدرجات حرية = ٣

(ثانيا) لكى تكون المقارنات مستقلة يلزم توفر ما يلى :

(١) شرط المقارنة (١٨) وهو $\sum ا_{ر} = 0$

(٢) شرط الاستقلال (٢١) وهو $\sum \frac{ا_{ر} ب_{ر}}{ن_{ر}} = 0$

مع ملاحظة أن المجموعات مختلفة الأحجام .

بقليل من العمليات الحسابية نجد أن الأوزان المطلوبة يمكن أن تؤخذ بالقيم المسجلة بالجدول الآتى أو بأى مجموعة من القيم تتناسب معها . ويمكن أن يسير العمل كالآتى :

جدول (٨ – ١٨)

المتوسط / المقارنة	(١) ٥	(٢) ٤	(٣) ٦	(٤) ٤
$\hat{\psi}_1$: (١) ضد (٢)	١	– ١	٠	٠
$\hat{\psi}_2$: (١،٢) ضد (٣)	٣	٤	– ٧	٠
$\hat{\psi}_3$: (١، ٢ ،٣) ضد (٤)	٣	٤	٥	– ١٢

نبدأ بأسهل المقارنات وهى $\hat{\psi}_1$ ونضع لها الوزنين ١، − ١ ثم نفرض أن أوزان المقارنة $\hat{\psi}_2$ هى ا_1 ، ا_2 ، ا_3 . لتحقيق شرط المقارنة ينبغى أن يكون $\text{ا}_1 + \text{ا}_2 + \text{ا}_3 = ٠$
ولتحقيق شرط استقلال $\hat{\psi}_1$ و $\hat{\psi}_2$ ينبغى أن يكون :

$$\frac{١ \times \text{ا}_1}{٣} + \frac{-١ \times \text{ا}_2}{٤} + ٠ = ٠$$

مع ملاحظة أن ٣ هو حجم العينة الأولى ، ٤ حجم العينه الثانية .

وإذن $\frac{\text{ا}_1}{٣} = \frac{\text{ا}_2}{٤}$ أى $\text{ا}_1 : \text{ا}_2 = ٣ : ٤$

فإذا أخذنا $\text{ا}_1 = ٣$ فإن $\text{ا}_2 = ٤$ ، $\text{ا}_3 = -٧$

وهذه هى الأعداد التى وضعت بالصف الثانى بالجدول .

بالمثل ، إذا فرضنا أن أوزان المقارنة $\hat{\psi}_3$ هى ب_1 ، ب_2 ، ب_3 ، ب_4 فإن
$\text{ب}_1 + \text{ب}_2 + \text{ب}_3 + \text{ب}_4 = ٠$

لاستقلال $\hat{\psi}_1$ ، $\hat{\psi}_3$: $\frac{١ \times \text{ب}_1}{٣} + \frac{-١ \times \text{ب}_2}{٤} = ٠$ ، ومنها $\text{ب}_1 : \text{ب}_2 = ٣ : ٤$

لاستقلال $\hat{\psi}_2$ ، $\hat{\psi}_3$: $\frac{٣ \, \text{ب}_1}{٣} + \frac{٤ \, \text{ب}_2}{٤} + \frac{-٧ \, \text{ب}_3}{٥} = ٠$

ومنها $\text{ب}_1 + \text{ب}_2 - \frac{٧}{٥} \text{ب}_3 = ٠$. فإذا أخذنا $\text{ب}_1 = ٣$ ، $\text{ب}_2 = ٤$ فإن
$\text{ب}_3 = ٥$ ، $\text{ب}_4 = -١٢$

وهذه هى الأعداد التى وضعت بالصف الثالث من الجدول .

(ثالثا) $\hat{\psi}_1 = ١ \times ٥ + (-١) \times ٤ = ١$

$\text{ع}^2(\hat{\psi}_1) = ١^2 / [\frac{١}{٣} \times ١^2 + \frac{١}{٤} \times (-١)^2] = ١ / \frac{٧}{١٢} = \frac{١٢}{٧} = ١,٧١٤$

$\therefore \text{ف}_{\text{ى}} = \frac{١,٧١٤}{٢,٤٦} < ١$ ليست ذات دلالة

كذلك

$\hat{\psi}_2 = ٣ \times ٥ + ٤ \times ٤ - ٧ \times ٦ = -١١$

مم $(\hat{\psi}_2) = (-١١)^٢ / (\frac{٩}{٣} + \frac{٦}{٤} + \frac{٤٩}{٥}) = ١٢١ / \frac{٤}{٥}١٦ = ٧,٢٠٢$

$\therefore$ $ف_ى = \frac{٧,٢٠٢}{٢,٤٦} = ٢,٩٢٨$ ليست ذات دلالة عند المستوى ٠,٠٥

لأن $ف_{٠,٠٥ [١ ، ١٣]} = ٤,٧٥$

بالمثل

$\hat{\psi}_3 = ١٣$ ، مم $(\hat{\psi}_3) = ٤,١٤٢$ ، $ف_ى = ١,٦٨٤$ ليست ذات دلالة

نلاحظ أن مم (المقارنات الثلاث) = ١,٧١٤+٧,٢٠٢ + ٤,١٤٢

= ١٣,٠٥٨ حيث $\nu = ٣$

وهذا يساوى م م (بين الأقسام) كما نتوقع .

جدول التباين هو :

جدول (٨ – ١٩)

مصدر التباين	م م	د ح	تقدير التباين	ف
بين الأقسام	١٣,٠٥٨	٣	٤,٣٥٣	١,٧٧
$\hat{\psi}_1$	١,٧١٤	١	١,٧١٤	١ >
$\hat{\psi}_2$	٧,٢٠٢	١	٧,٢٠٢	٢,٩٢٨
$\hat{\psi}_3$	٤,١٤٢	١	٤,١٤٢	١,٦٨٤
الخطأ	٣١,٩٨٠	١٣	٢,٤٦٠	
الكلى	٤٥,٠٣٨	١٦		

بما أن $ف_{٠,٠٥[٣ ، ١٢]} = ٥,٩٥$ ، $ف_{٠,٠٥[١ ، ١٢]} = ٤,٧٥$

نقبل الفرض الصفرى بأنه لا يوجد فروق بين أقسام المعالجة ، كما نقبل أن $\psi_١ = ٠$ ، $\psi_٢ = ٠$ ، $\psi_٣ = ٠$.

(٨ – ١٣) اختبار المقارنات البعدية :

نعلم أننا لا نجرى اختبارات المقارنات البعدية إلا إذا وجدنا من تحليل التباين أن هناك دلالة لعامل التجريب أوضحتها قيمة ف . كما نعلم أنه فى اختبار هذه المقارنات لا يهمنا أن تكون مستقلة كما هو الحال فى المقارنات القبلية .

وقد قدمنا بالبند (٨ – ٥ – ٢) أسلوبا لاختبار المقارنات البعدية يعتمد على إيجاد قيمة حرجة لمجموع المربعات إذا تعداها مجموع المربعات بين قسمين أو مجموعة من الأقسام يكون الاختلاف بينها ذا دلالة وإلا فهو ليس كذلك . ويتميز هذا الأسلوب بالبساطة والعمومية ويمكن استخدامه لاختبار أى مقارنة دون اشتراط تساوى أحجام العينات ، وهو فى الوقت نفسه غير حساس للانحرافات عن الاعتدالية وعدم تجانس التباينات . والصيغة التى استخدمناها لذلك هى المتباينة (٦) بالبند المذكور وهى :

$$م\ م\ (\text{بين الأقسام}) \geqslant (ك - ١)\ ع^٢_خ \cdot ف_{\alpha[ك - ١ ، ن - ك]} \qquad (٣٢)$$

وهذه المتباينة يمكن كتابتها بدلالة قيمة المقارنة $\hat{\psi}$ والخطأ المعيارى لها $ع_{\psi}$ كالآتى :

$$(\hat{\psi})^٢ \geqslant (ك - ١)\ ع^٢_{\psi} \cdot ف_{\alpha[ك - ١ ، ن - ك]} \qquad (٣٣)$$

حيث $ع^٢_{\psi} = ع^٢_خ\ مج \frac{أ^٢_ر}{ن_ر}$ ، ك عدد أقسام المعالجة ، ن الحجم الكلى للعينات .

والعدد الذى بالطرف الأيسر من المتباينة (٣٣) هو القيمة الحرجة للقمية $(\hat{\psi})^٢$. فإذا كانت $(\hat{\psi})^٢$ مساوية أو أكبر منها يرفض الفرض الصفرى $\psi = ٠$ عند المستوى α وتكون هذه المقارنة هى إحدى العوامل التى تسببت فى الدلالة العامة

لعامل التجريب ، أما إذا كانت $\hat{\psi}^{٢}$ أقل من القيمة الحرجة فإنها لا تكون ذات دلالة .

مثال (٨ – ١٥) :

فى المثال (٨ – ٥) كان أحد التساؤلات البعدية هو : هل متوسط السكروز يختلف عن متوسط السكريات الثلاثة الأخرى مجتمعة ؟

هذا التساؤل يمكن وضعه على هيئة مقارنة كالآتى :

$$\hat{\psi} = \frac{١}{٣}\bar{س}_{١} + \frac{١}{٣}\bar{س}_{٢} + \frac{١}{٣}\bar{س}_{٣} - \bar{س}_{٤}$$

$$= \frac{١}{٣}(٥٩,٣ + ٥٨,٢ + ٥٨) - ٦٤,١ = -٥,٦$$

$$(\hat{\psi})^{٢} = ٣١,٣٦$$

$$ع^{٢}_{\psi} = ٥,٤٦ \times \frac{١}{١٠}\left(\frac{١}{٩} + \frac{١}{٩} + \frac{١}{٩} + ١\right)$$
$$= ٥,٤٦ \times ٠,١٣٣٣ = ٠,٧٢٨$$

∴ القيمة الحرجة $= ٤ \times ٠,٧٢٨ \times ٢,٥٨ = ٧,٥١٣$

بما أن $٣١,٣٦ > ٧,٥١٣$ نرفض أن $\psi = ٠$ عند مستوى الدلالة ٠,٠٥

(٨ – ١٣ – ١) مقارنة أزواج المتوسطات :

تستخدم المتباينة (٣٣) للاختبارات البعدية لأزواج المتوسطات . ففى المثال (٨ – ١) كان لدينا خمسة أقسام واذن يكون هناك $^{ه}ق_{٢} = ١٠$ اختبارات كل

منها على الصورة $\bar{س}_ر - \bar{س}_ن$. فى هذه الحالة يكون لدينا قيمة حرجة واحدة لجميع الاختبارات لأن قيمة $ع^٢_ف$ واحدة لأى مقارنة وهى :

$$ع^٢_ف = ٥,٤٦ \times \frac{١}{١٠} (١+١) = ١,٠٩٢$$

وتكون القيمة الحرجة هى ٤ × ١,٠٩٢ × ٢,٥٨ = ١١,٢٦٩

ومن المناسب هنا وضع جميع المقارنات العشرة (الفروق بين المتوسطات) فى جدول كالآتى حيث الأعداد داخل الجدول تعبر عن الفروق بين المتوسطات .

جدول (٨ – ٢٠)

المتوسطات	(١) ٥٩,٣	(٢) ٥٨,٢	(٣) ٥٨,٠	(٤) ٦٤,١	(٥) ٧٠,١
(١) ٥٩,٣	٠	١,١	١,٣	-٤,٨*	-١٠,٨*
(٢) ٥٨,٢		٠	٠,٢	-٥,٩*	-١١,٩*
(٣) ٥٨,٠			٠	-٦,١*	-١٢,١*
(٤) ٦٤,١				٠	-٦,٠*
(٥) ٧٠,١					٠

وبمضاهاة القيمة المطلقة لكل من هذه الفروق بالجذر التربيعى للقيمة الحرجة وهو $\sqrt{١١,٢٦٩}$ = ٣,٣٥٧ نجد أن سبعة منها ذات دلالة عند المستوى ٠,٠٥ وهى المشار إليها بنجمة فى الجدول .

تمارين (٨ – ٦)

(١) فى احدى التجارب ذوات العامل الواحد كان هناك خمسة مستويات لعامل التجريب ، وقد اختير للتجريب خمس عينات عشوائية مستقلة بكل منها ١٠ مشاهدات وجد أن متوسطاتها كما هو مبين بالجدول الآتى :

المقارنات	المتوسطات (١)	(٢)	(٣)	(٤)	(٥)
	٨٦	٩٥	٩٢	٨٠	١٠٤
$\hat{\psi}_1$	١	١–	٠	٠	٠
$\hat{\psi}_2$	٠	٠	٠	١	١–
$\hat{\psi}_3$	$-\frac{1}{2}$	$-\frac{1}{2}$	٠	$\frac{1}{2}$	$\frac{1}{2}$
$\hat{\psi}_4$	$\frac{1}{4}$	$\frac{1}{4}$	١–	$\frac{1}{4}$	$\frac{1}{4}$

(أولا) اثبت أن المقارنات المدونة بالجدول مستقلة وأوجد قيمة كل منها .

(ثانيا) أوجد مجموع مربعات كل مقارنة واكتب جدول التباين بالتفصيل علما بأن المجموع الكلى للمربعات ٥١١٢ . بين أن هناك دلالة لعامل التجريب ثم ابحث دلالة كل من المقارنات الأربعة .

(ثالثا) أجر الاختبارات البعدية للفرق بين كل زوج من المتوسطات الخمسة .

(٢) أجريت تجربة لمعرفة العلاقة بين حجم ولون الحائط لحجرة تستخدم فى أسلوب مقنن للمقابلات interviews وبين مستوى القلق للمختبرين ، وجاءت النتائج كما فى الجدول الآتى :

الحجم \ اللون	أزرق	أخضر	أصفر	أحمر	المجموع
صغير	٨٦	١٠٤	١٣٤	١٦٠	
	٧١	١٧٥	١٣٩	١٥٥	
	١١٢	٩٦	١٤٤	١٧٠	
	٢٦٩	٣٧٥	٤١٧	٤٨٥	١٥٤٦
متوسط	١١٠	٨٣	١٥٠	١٧٥	
	٨٧	٨٩	١٥٦	١٥٢	
	١٠٠	٧٩	١٥٩	١٦٧	
	٢٩٧	٢٥١	٤٦٥	٤٩٤	١٥٠٧
مرتفع	١٠٥	٨٤	١٧٠	١٨٠	
	٩٣	٨٦	١٣٣	١٥٤	
	٨٥	٨٣	١٢٨	١٤١	
	٢٨٣	٢٥٣	٤٣١	٤٧٥	١٤٤٢
المجموع	٨٤٩	٨٧٩	١٣١٣	١٤٥٤	٤٤٩٥
المتوسط	٩٤,٣٣	٩٧,٦٧	١٤٥,٨٩	١٦١,٥٦	١٢٤,٨٦

(أولا) أجر تحليل التباين على أساس احتمال وجود تفاعل بين الخاصتين .

(ثانيا) أجر المقارنات القبلية الآتية بين الأعمدة (الألوان) واختبر دلالتها :

$\hat{\psi}_1 = \bar{\text{س}}_1 - \bar{\text{س}}_2 - \bar{\text{س}}_3 + \bar{\text{س}}_4$ ، $\hat{\psi}_2 = -\bar{\text{س}}_1 + \bar{\text{س}}_4$ ، $\hat{\psi}_3 = -\bar{\text{س}}_2 + \bar{\text{س}}_3$

(ثالثا) أجر المقارنات البعدية بين جميع الفروق بين أزواج متوسطات الأعمدة (الألوان) ،

(رابعا) اختر مقارنتين مستقلتين بين الصفوف (الأحجام) واختبر دلالة كل منهما .

(٨ – ١٤) النموذج عشوائي التأثيرات RANDOM EFFECTS MODEL

فى تناولنا لتحليل التباين للتجارب ذوات العامل الواحد بالبند (٨ – ٤) افترضنا أن لدينا مجتمعا معتدلا متوسطه μ وتباينه σ^2 أخذت منه عينة عشوائية قسمت عشوائيا إلى ك من المجموعات لتتلقى ك من المعالجات المختلفة مما قد يؤدى إلى اختلاف تراكيب هذه المجموعات ، ولذلك اعتبرنا أن هذه المجموعات هى عينات مأخوذة من مجتمعات (معتدلة) قد تختلف متوسطاتها μ_1 ، μ_2 ، ... ، μ_k وإن كنا نعتبر أن لها تباين مشترك σ^2 هو تباين المجتمع الأصلى . ونظرا لأن عناصر كل مجموعة تتلقى معالجة محددة فإننا نعتبر أن تأثير أى معالجة هو تأثير ثابت على جميع عناصر المجموعة التى تلقت هذه المعالجة ، ويختلف هذا التأثير من مجموعة إلى أخرى . ومن ثم وصفنا النموذج الإحصائى الذى استخدمناه فى التحليل بأنه نموذج ثابت التأثيرات أو النموذج I ووضعناه بالصيغة (٢) وهى :

س$_{رو}$ = μ + $\alpha_و$ + خ$_{رو}$ حيث ر = ١ ، ٢ ، ... ، ن$_و$ ؟ و = ١، ٢، ...، ك

وحيث س$_{رو}$ هو العنصر الرائى من المجموعة و التى تلقت المعالجة و

، $\alpha_و$ = $\mu_و$ − μ (الفرق بين متوسط المجتمع و ومتوسط المجتمع الأصلى) وهو فرق ثابت بالنسبة لعناصر المجموعة و ويختلف من مجموعة إلى أخرى .

، خ$_{رو}$ تعبر عن أخطاء التجريب وهو متغير عشوائى افترضنا أن له توزيع معتدل : مع (٠ ، σ) .

ويهدف هذا النموذج إلى المقارنة بين المتوسطات μ_1 ، μ_2 ، ... ، μ_k .

فى هذا النموذج ننظر إلى المعالجات التى استخدمت فى التجريب على أنها تستغرق جميع المعالجات ذات الأهمية وتقتصر استنتاجاتنا فقط على هذه المعالجات الممثلة فى التجربة . ولكن هناك أنواع من التجارب يكون المطلوب فيها التوصل إلى استنتاجات عن مجموعة كبيرة من المعالجات تشمل المعالجات الممثلة وغير الممثلة فى التجربة . أى أن الباحث يكون مهتما بمجموعة كبيرة من المعالجات الممكنة لعامل التجريب ولكنه حين يقوم بالتجربة لا يأخذها جميعها بل يأخذ عينة عشوائية منها ومع ذلك يرغب فى التوصل إلى استنتاجات عن تأثيرات المجموعة الكاملة . فى هذه الحالة يستخدم نموذجا إحصائيا آخر يسمى بالنموذج عشوائى التأثيرات أو بالنموذج II يصاغ كالآتى :

$$س_{ر و} = \mu + أ_{و} + خ_{ر و} \qquad \text{حيث } ر = ١، ٢ ، ...، ن_{و} ؟$$
$$و = ١، ٢، ...، ك \qquad (٣٤)$$

وحيث $أ_{و} = \mu(و) - \mu$

= الفرق بين المتوسط $\mu(و)$ للمعالجة و ومتوسط المجتمع الأصلى .

ولما كان $\mu(و)$ هو متغير عشوائى لأن المعالجة و تختار عشوائيا ، فإن $أ_{و}$ يكون بالضرورة متغيرا عشوائيا هو الآخر .

وهذا النموذج يشبه النموذج I إلا أن الفرق بينهما كبير ، فبينما نعتبر أن تأثيرات $\alpha_{و}$ فى النموذج I ثابتة ، فإننا نعتبر أن تأثيرات $أ_{و}$ عشوائية . أى أننا حين نكرر سحب العينات فإننا تحت النموذج I نسحب دائما نفس المعالجات بنفس التأثيرات $\alpha_{و}$ ، أما تحت النموذج II فنسحب فى كل مرة عينة عشوائية جديدة تختلف فيها تأثيرات $أ_{و}$. ومن ثم نصف تأثيرات $أ_{و}$ بأنها عشوائية وتتوقف على العينة المختارة . وكما سبق القول يستخدم النموذج I حين يكون المطلوب التوصل إلى استنتاجات عن فئة محددة من المعالجات تستخدم جميعها فى التجريب ، أما النموذج II فيستخدم حين تكون هناك فئة كبيرة من المعالجات التى تهم الباحث ولكنه لا يستخدمها جميعا بل يستخدم عينة عشوائية منها .

وفى النموذج II نفترض أن للمتغيرين العشوائيين اى وخـرى التوزيعين الآتيين :

$$\text{اى : مع } (\cdot ، \sigma)\text{، خـرى : مع } (\cdot ، \sigma) \qquad (٣٥)$$

كما نفترض أن قيم اى ، خـرى مستقلة عن بعضها البعض .

ويلاحظ أننا عرفنا اى بانها انحرافات متوسطات المعالجات عن المتوسط العام للمجتمع ولذلك فإن متوسط تأثيرات [illegible] يساوى صفرا ، وهذا ما دعانا لأن نفرض أن متوسط توزيع اى صفر ، أما تباين هذا التوزيع فهو مقدار مجهول $\sigma^{٢}$ يراد تقديره وتقدير مدى مساهمته فى الاختلاف الذى يظهر عند تحليل التباين وهذا أمر هام فى كثير من التطبيقات الإحصائية .

إن تحليل التباين يتخذ نفس الأسلوب الحسابى فى النموذجين I وII إلا أن الهدف من التحليل يختلف تماما ، فبينما يهدف التحليل فى النموذج I إلى مقارنة المتوسطات ، فهو يهدف فى النموذج II إلى دراسة التباينات ، وبصفة خاصة إلى تقدير واختبار كل من التباينين $\sigma^{٢}$ $\sigma^{٢}$ وكذلك إلى تقدير التباين $\bar{\sigma}^{٢}_{\text{س}}$ لتوزيع المعاينة للمتوسط الحسابى لأهميته فى تقدير مدى الثقة فى تقدير المتوسط الحقيقى μ للمجتمع عن طريق المتوسط $\bar{\text{س}}$ للعينات.أما متوسطات تأثيرات اى فلا جدوى من محاولة تقديرها أو تقدير الفروق بينها لأنها عشوائية .

إذا رمزنا بالرمز $ع^{٢}_{ع}$ للتباين المشاهد داخل أقسام المعالجة : ط م (داخل الأقسام) ، وبالرمز $ع^{٢}_{ب}$ للتباين المشاهد بين الأقسام : ط م (بين الأقسام) وبالرمز ن لعدد قيم المتغير فى أى قسم فيمكن إثبات ما يلى :

(١) $ع^{٢}_{ع}$ هو تقدير غير متحيز للتباين $\sigma^{٢}$ (٣٦)

(٢) $ع^2_ب$ هو تقدير غير متحيز للمقدار $\sigma^2 + ن\,\sigma$. (٣٧)

وإذا رمزنا بالرمز $ع^2_1$ للتقدير غير المتحيز للتباين σ^2 فإن

$$ع^2_1 = \frac{1}{ن}(ع^2_ب - ع^2_خ) \qquad (٣٨)$$

وهذه نتيجة مباشرة من (١) ، (٢) ، كما ينتج أن الفرض الصفرى $\sigma^2_1 = ٠$ يمكن أن يختبر بواسطة اختبار ف بالصورة الآتية لأنه فى حالة صحة هذا الفرض يكون كل من $ع^2_ب$ ، $ع^2_خ$ تقديرا غير متحيز للتباين σ^2 :

$$ف = \frac{ع^2_ب}{ع^2_خ} \quad \text{بدرجتى حرية } (ك - ١ ، هـ - ك) \qquad (٣٩)$$

حيث ك عدد أقسام المعالجة و هـ = الحجم الكلى للعينات = ك ن .

(٣) تقدير σ^2 هو $\frac{ع^2_ب}{هـ}$ (٤٠)

أى أن التباين $\sigma^2_{\bar{س}}$ لتوزيع المعاينة للمتوسط الحسابى يقدر بواسطة التباين المشاهد بين الأقسام مقسوما على العدد الكلى للمشاهدات هـ .

تتبين نوعية التجارب التى تستلزم النموذج II من المثالين الآتيين .

مثال (٨ – ٢٦) :

فى دراسة لمحتوى الكلسيوم فى أوراق نبات اللفت الأخضر أخذت عينة عشوائية من ٤ أوراق من هذا النبات ثم أخذ من كل ورقة اختيرت ٤ أجزاء وزن كل منها ١٠٠ جرام وبذلك تجمع ١٦ جزءا من أوراق النبات . قيست النسبة المئوية لمحتوى الكلسيوم فى كل منها وسجلت القياسات بالجدول (٨ – ٢١) الآتى :

جدول (٨ – ٢١)

النسب المئوية للكلسيوم فى أوراق نبات اللفت : ك = ٤ أوراق ، ن = ٤ أجزاء .

	الأوراق			
	(١)	(٢)	(٣)	(٤)
النسبة المئوية للكلسيوم في أجزاء الورقة	٣,٢٨	٣,٥٢	٢,٨٨	٣,٣٤
	٣,٠٩	٣,٤٨	٢,٨٠	٣,٣٨
	٣,٠٣	٣,٣٨	٢,٨١	٣,٢٣
	٣,٠٣	٣,٣٨	٢,٧٦	٣,٢٦
المجموع للورقة	١٢,٤٣	١٣,٧٦	١١,٢٥	١٣,٢١
المتوسط للورقة	٣,١١	٣,٤٤	٢,٨١	٣,٣٠

م = ٥٠,٦٥

$\bar{س}$ = ٣,١٧

نظرا لأن الأوراق تختلف من جوانب كثيرة قد لا نعرف طبيعتها كالاختلافات الوراثية والبيئية ونظرا لأننا اخترنا عينة عشوائية منها فإن النموذج المناسب هو النموذج عشوائى التأثير بالصيغة (٣٤) حيث $س_{رو}$ ترمز إلى نسبة الكلسيوم فى الجزء ر من الورقة و ، ر = ١ ، ٢ ، ٣ ، ٤ ؟ و = ١ ، ٢ ، ٣ ، ٤ وحيث μ ترمز إلى المتوسط الحقيقى لنسبة الكلسيوم فى مجتمع أوراق اللفت الأخضر . كما أن :

σ_1^2 ترمز إلى التباين بين **الأوراق** أى إلى مدى أثر الاختلافات الوراثية بين ورقة وأخرى على محتوى الكلسيوم ،

σ^2 ترمز إلى التباين **داخل الأوراق** أى بين القياسات داخل كل ورقة وبالتالى فهى تعبر عن تأثير الاختلافات غير الوراثية على محتوى الكلسيوم .

المطلوب فى هذا المثال ما يلى :

(أولا) اختبار وجود أو عدم وجود تأثيرات ترجع إلى المعالجات أى إلى العوامل الوراثية . ويتضمن هذا الاختبار اختبار الفروض الصفرية : $أ_١ = ٠$ ، $أ_٢ = ٠$ ، $أ_٣ = ٠$ ، $أ_٤ = ٠$. وتتحقق هذه الفروض إذا وفقط إذا تحقق الفرض $\sigma_١^2 = ٠$ ولذلك فإن الفرض الصفرى المطلوب اختباره يؤول إلى الآتى :

$$ف_. : \sigma_١^2 = ٠$$

وهذا الاختبار لا يتعلق فقط بوجود تأثيرات للمعالجات التى استخدمت فى التجربة بل لجميع المعالجات الممكنة التى دخلت والتى لم تدخل فيها .

(ثانيا) تقدير المتوسط μ لمحتوى الكلسيوم فى مجتمع أوراق اللفت مع تقدير درجة الثقة فى هذا التقدير . وهذا هو الهدف الرئيسى فى هذه التجربة .

(ثالثا) المقارنة بين التباينين $\sigma_١^2$ و σ^2 لأن هذه المقارنة قد تشير إلى ضرورة تعديل التجربة وتصميم تجربة مماثلة تكون أكثر دقة وأقل تكلفة . فإذا كان التباين $\sigma_١^2$ بين الأوراق أكبر نسبيا من التباين σ^2 داخل الأوراق فالأفضل تصميم تجربة نأخذ فيها عددا أكبر من الأوراق وعددا أصغر من الأجزاء والعكس بالعكس ، لأن هذا من شأنه تصغير تباين توزيع المعاينة للوسط الحسابى فيكون تقديرنا للبارامتر μ أكثر دقة .

نقوم الآن بتحليل التباين للمثال (٨ – ١٦) .

$$\text{عامل التصحيح} = \frac{م^2}{ن} = \frac{٢٥٠,٦٥}{١٦} = ١٦٠,٣٣٨٩$$

مم (الكلى) $= ٣,٢٨^٢ + ٣,٠٩^٢ + ... + ٣,٢٦^٢ - ١٦٠,٣٣٨٩$

$= ٠,٩٦٦٦$ بدرجات حرية ١٥

مم (بين الأوراق) $= \frac{١}{٤}(١٢,٤٣^٢ + ١٣,٧٦^٢ + ١١,٢٥ + ١٣,٢١^٢)$

$- ١٦٠,٣٣٨٩$

$= ٠,٨٨٨٣٧$ بدرجات حرية ٣

مم (داخل الأوراق) $= ٠,٩٦٦٦ - ٠,٨٨٨٣٧$

$= ٠,٠٧٨٢٣$ بدرجات حرية ١٢

جدول (٨ - ٢٢)

مصدر التباين	م م	د ح	ط م	التباين المتوقع
بين الأوراق	٠,٨٨٨٣٧	٣	$ع^٢_ب = ٠,٢٩٦١$	$\sigma^٢ + ٤\sigma^٢_١$
بين القياسات داخل الأوراق	٠,٠٧٨٢٣	١٢	$ع^٢_ع = ٠,٠٠٦٦$	$\sigma^٢$
الكلى	٠,٩٦٦٦	١٥		

(أولا) : الفرض الصفرى : $\sigma^٢_١ = ٠$ (لا يوجد تأثير للخواص الوراثية على محتوى الكلسيوم)

من (٣٩) : $ف_ى = \frac{٠,٢٩٦١}{٠,٠٠٦٦} = ٤٤,٩^{**}$

وهذه القيمة تزيد كثيرا عن القيمة الحرجة $ف_{٠,٠١[٣،١٢]}$ = ٥,٩٥ مما يجعلنا نرفض الفرض الصفرى عند مستوى عالى من الدلالة ونستنتج أن هناك تفاوتا كبيرا فى الخواص الوراثية لأوراق النبات يؤثر تأثيرا فعالا فى محتوى الكلسيوم فى هذه الأوراق .

(ثانيا) : نقدر الوسط الحسابى μ للنسبة المئوية لمحتوى الكلسيوم فى مجتمع أوراق نبات اللفت الأخضر بواسطة الوسط الحسابى العام للعينات وهو = ٣,١٧ ولبيان مدى الدقة فى هذا التقدير وحساب فترات الثقة للمتوسط μ نستخدم الصيغة (٤٠) لتقدير تباين توزيع المعاينة للوسط الحسابى كالآتى :

$$\text{تقدير } \sigma^2_{\bar{س}} = \frac{ع^2_ب}{ه} = \frac{٠,٢٩٦١}{١٦} = ٠,٠١٨٥ \qquad \text{بدرجات حرية ٣}$$

ويمكننا أن نوجد فترة ثقة للمتوسط μ بدرجة ٩٩٪ كالآتى :

$$\text{الحد الأدنى للفترة} = ٣,١٧ - \sqrt{٠,٠١٨٥} \times ت_{٠,٠١\,[٣]}$$

$$= ٣,١٧ - ٠,١٣٦٠ \times ٥,٨٤١ = ٢,٣٧٦$$

$$\text{الحد الأعلى للفترة} = ٣,١٧ + ٠,١٣٦٠ \times ٥,٨٤١ = ٣,٩٦٤$$

أى أن فترة الثقة المطلوبة هى (٢,٣٧٦ ، ٣,٩٦٤) .

(ثالثا) : من (٣٦) ، التقدير غير المتحيز للتباين σ^2 هو $ع^2_ع = ٠,٠٠٦٦$

ومن (٣٨) ، التقدير غير المتحيز للتباين σ^2_1 هو

$$ع^2_1 = \frac{1}{ن}(ع^2_ب - ع^2_ع) = \frac{1}{٤}(٠,٢٩٦١ - ٠,٠٠٦٦)$$
$$= ٠,٠٧٢٤$$

$$\text{نلاحظ أن } \frac{ع^2_1}{ع^2_ع} = \frac{٠,٠٧٢٤}{٠,٠٠٦٦} = ١٠,٩٧$$

أى أن تقدير σ^2 عشرة أضعاف أو أكثر تقدير σ^2 ولو أردنا تحسين التجربة وتحسين الخطأ المعيارى لتوزيع المعاينة للوسط الحسابى ينبغى أن نزيد من عدد الأوراق وأن نقلل من عدد الأجزاء التى تؤخذ من كل ورقة .

ملاحظة :

فى هذا المثال اخترنا عينة عشوائية من الوحدات (أوراق النبات) التى يمكن أن نسميها بالوحدات الإبتدائية للمعاينة primary sampling units ثم اخترنا عينات عشوائية جزئية من كل وحدة من الوحدات الابتدائية ويمكن أن نستمر فى ذلك بأخذ عينات عشوائية من كل عنصر من عناصر العينات الجزئية السابق اختيارها وأن نكرر ذلك حسبما تقتضى التجربة . إن مثل هذه المعاينة تسمى بالمعاينة ذات المراحل أو بالمعاينة العشية multistage or nested sampling حيث تحدث عدة تقسيمات متدرجة ومتداخلة كأعشاش الطيور . وسوف نعود إلى ذلك فى البند (١٥ - ٥) من الفصل الخامس عشر .

مثال (٨ - ١٧) :

فى إحدى التجارب النفسية كان يشك فى أن شخصية المجرب (القائم بالتجريب) لها تأثير على النتائج التى يتوصل إليها . ونظرا لأن هناك عددا كبيرا من المجربين الذين يمكنهم القيام بالتجربة مما يعوق استخدامهم جميعا فقد اختير منهم عينة عشوائية من خمسة مجربين لإجراء التجربة تحت نفس الظروف على أن يستخدم كل منهم مجموعة من ٨ أشخاص تختار عشوائيا وعلى أن توزع المجموعات على المجربين عشوائيا . سجلت البيانات الناتجة من التجارب الخمس فى الجدول (٨ - ٢٣) الآتى ، والمطلوب بحث ما إذا كان لشخصيات المجربين أثر على نتائج التجربة .

الجدول (٨ - ٢٣)

المجربون					
(١)	(٢)	(٣)	(٤)	(٥)	
٥,٨	٦,٠	٦,٣	٦,٤	٥,٧	
٥,٩	٦,١	٥,٥	٦,٤	٥,٩	
٥,٧	٦,٦	٥,٧	٦,٥	٦,٥	
٥,٩	٦,٥	٦,٠	٦,١	٦,٣	
٥,٦	٥,٩	٦,١	٦,٦	٦,٢	
٥,٤	٥,٩	٦,٢	٥,٩	٦,٤	
٥,٣	٦,٤	٥,٨	٦,٧	٦,٠	
٥,٢	٦,٣	٥,٦	٦,٠	٦,٣	
٤٤,٠	٥٠,٦	٤٧,٢	٥٠,٦	٤٩,٣	٢٤٠,٨

الحل :

نظرا لأن هناك عددا كبيرا من الأفراد الذين يمكنهم القيام بالتجربة ، لكل منهم شخصية تميزه عن الآخرين ، ونظرا لأننا اخترنا عينة عشوائية منهم ، فإن النموذج المناسب لهذه التجربة هو النموذج عشوائى التأثير بالصيغة (٣٤) ، ونعتبر أن لدينا خمس معالجات يمثلها خمسة مجربين .

عامل التصحيح $= \frac{(٢٤٠,٨)^2}{٤٠} = ١٤٤٩,٦١٦$

م٢ (الكلى) $= ٥,٨^٢ + ٥,٩^٢ + ... + ٦,٠^٢ + ٦,٣^٢$
$- ١٤٩٩,٦١٦$

$= ٦,٣٢$ بدرجات حرية ٣٩

م٢ (بين المجربين) $= \frac{١}{٨}(٤٤,٠^٢ + ٥٠,٦^٢ + ... + ٤٩,٣^٢) - ١٤٤٩,٦١٦$

$= ٣,٤٧$ بدرجات حرية ٤

م ٢ (داخل المجربين) $= ٦,٣٢ - ٣,٤٧ = ٢,٨٥$ بدرجات حرية ٣٥

نلخص هذه النتائج فى جدول التباين الآتي :

جدول (٨ — ٢٤)

مصدر التباين	م ٢	د ح	ط ٢	التباين المتوقع	$ف_ى$
بين المجربين	٣,٤٧	٤	$ع^٢_ب = ٠,٨٦٨$	$\sigma^2 + ٨\sigma^2_١$	$١٠,٧٢^{**}$
داخل المجربين	٢,٨٥	٣٥	$ع^٢_خ = ٠,٠٨١$	σ^2	
الكلى	٦,٣٢	٣٩			

الفرض الصفرى : $\sigma^2_١ = ٠$ (لا يوجد تأثير للمجربين على نتائج التجربة)

$ف_ى = \frac{٠,٨٦٨}{٠,٠٨١} = ١٠,٧٢^{**}$

وهذه القيمة تزيد عن القيمة الحرجة ف$_{٠,٠١[٤،٣٥]}$ التى تقع بين ٣,٨٣ ، ٤,٠٢ وإذن نرفض الفرض الصفرى عند مستوى الدلالة ٠,٠١ ونحكم بأن هناك دليلا كافيا على صحة القول بأن لشخصيات المجربين تأثيرا فيما يتوصلون إليه من نتائج فى هذه التجربة . وهذه نتيجة خطيرة ينبغى أن تؤخذ فى الاعتبار عند تفسير نتائج التجربة . وتتبين هذه الخطورة أيضا إذا حسبنا النسبة التى ساهم بها التباين σ^2_1 من التباين الكلى (وهو ٦,٣٢ ÷ ٣٩ = ٠,١٧٩) :

من (٣٨) : التقدير غير المتحيز للتباين σ^2_1 هو

$$ع^2_1 = \frac{1}{8} (٠,٨٦٨ - ٠,٠٨١) = ٠,٠٩٨$$

∴ النسبة التى ساهم بها التباين σ^2_1 من التباين الكلى = $\frac{٠,٠٩٨}{٠,١٧٩}$ ≑ ٠,٥٥ وهى نسبة كبيرة تشير إلى أن أكثر من نصف التباين الكلى يرجع إلى تأثير اختلاف شخصيات المجربين .

ملاحظة :

يمتد استخدام النموذج عشوائى التأثيرات لتحليل التباين للتجارب ذوات العاملين أو أكثر .

الفصل التاسع

الانحدار الخطى البسيط

SIMPLE LINEAR REGRESSION

(٩ – ١) المجتمع ذو المتغيرين : BIVARIATE POPULATION

المجتمع ذو المتغيرين هو مجتمع ننظر إليه من حيث متغيرين س ، ص يكونان موضع اهتمامنا في وقت ما ، فمثلا في مجتمع من الطلاب قد نهتم بالمتغير س الذى يعبر عن نسبة الذكاء والمتغير ص الذى يعبر عن درجة الطالب في مادة الإحصاء وفي مجتمع من القمح قد تعبر س عن تاريخ الزراعة وتعبر ص عن مقدار المحصول الناتج ، وفي مجتمع من الأبقار قد تعبر س عن مقدار الغذاء اليومى وتعبر ص عن الزيادة في الوزن بعد مدة من الزمن .

كما سبق الذكر ، يميز الإحصاء بين نوعين من المتغيرات : المتغيرات العشوائية والمتغيرات غير العشوائية ، والمتغير العشوائي هو متغير حقيقي يخضع لمؤثرات عشوائية غير منظورة ولذلك لا نستطيع التحكم فيه تجريبياً . وهذا النوع من المتغير يكون له توزيع احتمال بمعني أنه لو كان من النوع الوثاب مثلا فإن ظهور أى عنصر من عناصره يكون مصحوباً باحتمال ما ، أما المتغير غير العشوائي (أو الرياضي) فليس له توزيع احتمال ويمكن التحكم فيه تجريبياً أو تحديد قيمه مقدماً أو قياسها بدقة أو بخطأ يمكن إهماله .

وفي دراستنا لمجتمع ذى متغيرين س ، ص كثيراً ما يكون اهتمامنا منصباً على إيجاد العلاقة بينهما إن كان هناك ثمة علاقة ، وعلى محاولة وضع هذه العلاقة على

هيئة دالة ص = د (س) . وتختلف هذه الدالة بطبيعة الحال باختلاف العلاقة بين المتغيرين فقد تكون على صورة خطية ص = $\alpha + \beta$ س أو على صورة تربيعية ص = $\alpha + \beta$ س + γ س^{2} أو على صورة أسية ص = α هـ$^{\beta س}$ وهكذا ... إلا أننا سوف نقصر اهتمامنا هنا على الحالات التي تكون فيها الدالة على الصورة الخطية .

سنبني هذه الدراسة على الافتراضين الآتيين وافتراض ثالث نقدمه في البند (٩ — ٥) ، والنموذج الذى سنستخدمه فى هذه الدراسة يسمى بالنموذج I لتحليل الانحدار .

الافتراض الأول :

« المتغير س هو متغير رياضي والمتغير ص هو متغير عشوائي » .

وهذا الافتراض يعني من الناحية التطبيقية أننا نبدأ بتحديد قيم ثابتة س$_{١}$ ، س$_{٢}$ ، ... ، س$_{ن}$ من المتغير س ثم نقوم بملاحظة أو قياس القيم (العشوائية) المناظرة ص$_{١}$ ، ص$_{٢}$ ، ... ، ص$_{ن}$ من المتغير ص . فمثلا قد تكون القيم السينية هي درجات حرارة معينة ١٠ ، ١٥ ، ٢٠ ، ... ، ٥٠ وتكون القيم الصادية هي مقادير ما نشاهده من تمدد معدن عند هذه الدرجات . أو تكون قيم س هى أطوال أطفال عند الولادة بالسنتمترات وتكون قيم ص هي مدد الحمل بالأيام . أو تكون قيم س هى أعماق محددة تحت سطح البحر وتكون القيم الصادية هى نسب الملوحة عند هذه الأعماق .

الافتراض الثانى :

« العلاقة بين المتغيرين س ، ص (في المجتمع) هى علاقة خطية » .

ونعني بهذا الفرض رياضياً أن يكون متوسط قيمة ص عند قيمة معينة س يأخذ الصيغة الخطية الآتية :

$$\mu_{س} = \alpha + \beta س \quad (١)$$

حيث α ، β بارامتران مجهولان ينبغي تقديرهما من العينة . هذا مع ملاحظة أنه عند أي قيمة س يمكن أن تأخذ ص قيماً كثيرة لأنها كما ذكرنا تتأثر بعوامل عشوائية لا نستطيع التحكم فيها ولهذا يعبر الطرف الأيمن من الصيغة (١) عن متوسط هذه القيم وهو بالطبع أحسن قيمة تناظر القيمة س . ويسمى المتغير س بالمتغير المستقل كما يسمى المتغير ص بالمتغير التابع .

إن هذا الافتراض لا يوضع إلا في الحالات التي نعلم بخبرتنا السابقة أو من معلومات خاصة أنه صحيح . ومع ذلك إذا كنا نشك في صحته فهناك طريقة إحصائية سنذكرها بعد للتحقق من سلامته .

إن مهمتنا الابتدائية في دراسة العلاقة الخطية بين متغيرين س ، ص هي إيجاد أحسن تقديرين للبارامترين المجهولين α ، β الموجودين بالعلاقة (١) ، ونعتمد في ذلك كالمعتاد على تصميم تجربة على عينة من المجتمع الذى ندرسه ومنها نحصل على قيم تجريبية للمتغيرين س ، ص تبدو كما يلي :

س١ س٢ س٣ ... سن

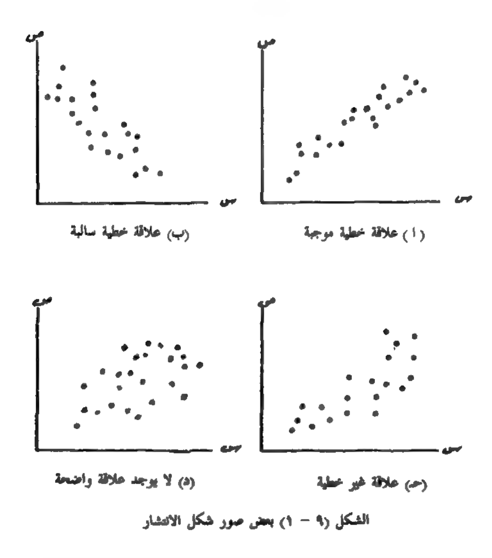
ص١ ص٢ ص٣ ... صن

ومن هذه الأزواج من القيم نوجد التقديرين المطلوبين كما سنرى بعد ، ويساعدنا التمثيل البياني لهذه الأزواج على تصور المشكلة التي نتناولها والخط الذى ننشده .

(٩ – ٢) شكل الانتشار : SCATTERGRAM

على أساس أن المتغيرين س ، ص حقيقيان نستطيع تمثيل أزواج القيم (سر ، صر) التى حصلنا عليها من العينة على نظام إحداثيات ذى محورين متعامدين . فيمثل كل زوج منها بنقطة معينة في المستوى ، ومجموعة النقط الناتجة تؤلف شكلا يسمى بشكل الانتشار . وإذا كان هناك ثمة علاقة بين المتغيرين فإن هذه النقط لا تكون مبعثرة كيفما اتفق بل تتخذ نمطاً معيناً يوحى بوجود وطبيعة هذه

العلاقة . فإذا بدا شكل الانتشار كما في أى من الشكلين (٩ – ١ – ا) أو (٩ – ١ – ب) فإنه يشير بصفة مبدئية بأن العلاقة بين المتغيرين هى علاقة خطية ، لأننا نستطيع أن نتصور وجود خط مستقيم تقع النقط من حوله وقريبة منه وإن كانت لا تمر جميعها به .



(ا) علاقة خطية موجبة

(ب) علاقة خطية سالبة

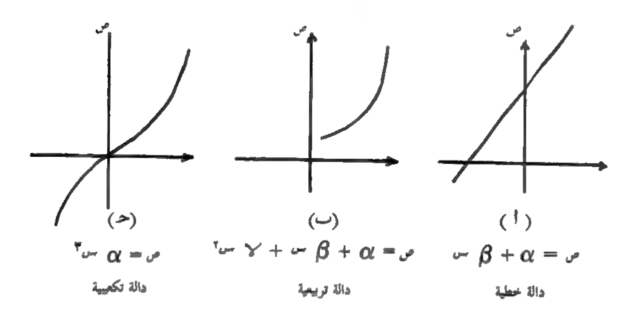
(حـ) علاقة غير خطية

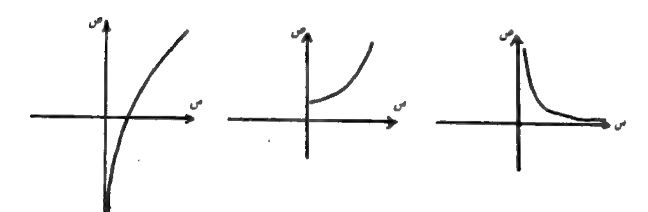
(د) لا يوجد علاقة واضحة

الشكل (٩ – ١) بعض صور شكل الانتشار

أما إذا بدا شكل الانتشار كما فى أى من الشكلين (حـ) ، (د) فإننا نشك فى خطية العلاقة بين المتغيرين وعلينا أن نبحث عن افتراض آخر غير افتراض الخطية نتعامل به مع المتغيرين . إن معرفة الباحث بخواص المنحنيات تساعده على كشف طبيعة

العلاقة بين المتغيرين واقتراح المعادلة التى تناسبها . وفيما يلى بعض المنحنيات التى تعاون على كشف الأنماط التى قد يشير إليها شكل الانتشار والدوال التى تعبر عنها جبريا . وسنرى فى البند (٩ — ٧) أنه يمكن تناول بعض هذه الدوال كدوال خطية بعد إجراء تحويلات مناسبة عليها .





الشكل (٩ — ٢) : بعض أنماط شكل الانتشار

(٩ – ٣) تقدير البارامترين α ، β .

نفرض أن ($س_ر$ ، $ص_ر$) هى إحدى القيم المشاهدة في العينة . حسب الافتراض الثاني ينبغى أن تكون العلاقة بين $س_ر$ ، $ص_ر$ على الصورة :

$$ص_ر = \alpha + \beta\, س_ر + خ_ر \qquad (٢)$$

حيث $خ_ر$ هو الفرق بين القيمة المشاهدة $ص_ر$ المناظرة للقيمة $س_ر$ وبين القيمة المتوقعة لها وهى $\alpha + \beta\, س_ر$ ولذلك تسمى القيم $خ_ر$ بالبواقي residuals وهى تعبر عن الأخطاء العشوائية في قياس المتغير العشوائي ص . ونظراً لأن هذه الأخطاء تكون بالزيادة لبعض قيم ص وبالنقصان للبعض الآخر ، فإننا نفترض أن متوسط هذه الأخطاء يؤول إلى الصفر على المدى البعيد .

طريقة المربعات الصغرى : LEAST SQUARES METHOD

كما أشرنا من قبل ، إن مهمتنا الأساسية هى استخدام أزواج القيم ($س_ر$ ، $ص_ر$) المشاهدة في العينة لإيجاد أحسن تقديرين للبارامترين α ، β . فإذا كان ا ، ب هما هذان التقديران فإن المعادلة :

$$\hat{ص} = ا + ب\, س \qquad (٣)$$

تكون معبرة أحسن تعبير عن المعادلة (١) ويكون الخط الممثل للمعادلة (٣) هو أحسن خط يلائم مجموعة نقط العينة . فإذا وفقنا في إيجاد ا ، ب سمى الخط (٣) بخط أحسن مطابقة line of best fit أو خط انحدار ص على س regression line of Y on X وأحياناً يسمى بصيغة التنبؤ prediction formula .

والطريقة الأكثر شيوعاً لإيجاد ا ، ب تبني على مبدأ يعرف بمبدأ المربعات الصغرى يمكن صياغته كالآتي ، مع ملاحظة الافتراض الأول عن أن قيم س ثابتة وقيم ص متغيرة .

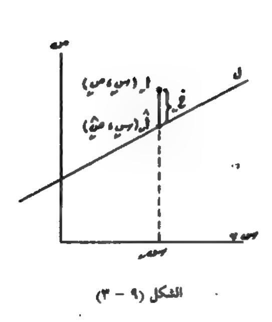
« أحسن خط يلائم مجموعة النقط ($س_ر$ ، $ص_ر$) هو ذلك الذى نحدده بحيث يكون مجموع مربعات البواقي $خ_ر$ أصغر ما يمكن » .

من المعادلة (٢) نرى أن $خ_ر = ص_ر - \alpha - \beta\, س_ر$

$$\text{ليكن } د\,(\alpha ، \beta) = \sum خ_ر^2 = \sum (ص_ر - \alpha - \beta\, س_ر)^2 \qquad (٤)$$

فإذا نظرنا للدالة (٤) على أنها دالة في المجهولين α ، β فإن مبدأ المربعات الصغرى يقول إن أحسن خط هو ذلك الذى تحدد فيه قيمتا α ، β بحيث تكون قيمة هذه الدالة نهاية صغرى .

ويوضح هذا المبدأ هندسياً كالآتي :
لتكن $ا_ر$ ($س_ر$ ، $ص_ر$) إحدى النقط المشاهدة في العينة أى إحدى نقط شكل الانتشار ، ولتكن $\hat{ا}_ر$ ($س_ر$ ، $\hat{ص}_ر$) هى نقطة واقعة على خط مستقيم ل وتشترك مع النقطة $ا_ر$ في الإحداثي السيني $س_ر$. إن مبدأ المربعات الصغرى يقول إن الخط ل يكون هو خط أحسن مطابقة إذا كان مجموع مربعات المسافات الرأسية بين النقط $ا_ر$ ، $\hat{ا}_ر$ نهاية صغرى ، أى إذا كانت الدالة



الشكل (٩ – ٣)

$$د\,(\alpha ، \beta) = \sum (ص_ر - \hat{ص}_ر)^2 = \sum (ص_ر - \alpha - \beta\, س_ر)^2$$

نهاية صغرى . بطرق التفاضل المعتادة نفاضل الدالة (٤) جزئياً بالنسبة إلى كل من α ، β . وبمساواة كل من الناتجين بالصفر نحصل على معادلتين آنيتين يكون

حلهما معاً هما القيمتان ا ، ب المطلوبتان. وهاتان المعادلتان تسميان بالمعادلتين المعتادتين وتأخذان الصورتين الآتيتين :

$$\text{مجـ ص} = ن\,\alpha + \beta\,\text{مجـ س}\ ،\ \text{مجـ س ص} = \alpha\,\text{مجـ س} + \beta\,\text{مجـ س}^2$$

وبحلهما معا ينتج ما يلى :

أحسن تقديرين للبارامترين α ، β من العينة هما ا ، ب حيث :

$$ب = \frac{ن\ \text{مجـ س ص} - (\text{مجـ س})\ (\text{مجـ ص})}{ن\ \text{مجـ س}^2 - (\text{مجـ س})^2} \qquad (٥)$$

وحيث ن هى عدد أزواج القيم ($س_ر$ ، $ص_ر$)

$$،\ ا = \bar{ص} - ب\,\bar{س} \qquad (٦)$$

وبالتالى تكون معادلة انحدار ص على س هى

$$\hat{ص} = (\bar{ص} - ب\,\bar{س}) + ب\,س$$

$$\text{أو}\ \hat{ص} = \bar{ص} + ب\,(س - \bar{س}) \qquad (٧)$$

حيث ب معطاة بالصيغة (٥) وتسمى بمعامل انحدار ص على س وهى تعبر عن ميل خط الانحدار (٧) عن الاتجاه الموجب لمحور السينات .

ملاحظة (١) :

من الواضح أن خط الانحدار (٧) يمر بالنقطة ($\bar{س}$ ، $\bar{ص}$) .

ملاحظة (٢) :

بقسمة كل من بسط ومقام الطرف الأيسر من (٥) على (ن − ١) وبعد عمليات جبرية بسيطة نجد أن :

$$\text{ب} = \frac{\frac{1}{\text{ن}-1}\ \text{مج}\ (\text{س} - \bar{\text{س}})\ (\text{ص} - \bar{\text{ص}})}{\frac{1}{\text{ن}-1}\ \text{مج}\ (\text{س} - \bar{\text{س}})^{2}} = \frac{\text{ع}_{\text{سص}}}{\text{ع}^{2}_{\text{س}}} = \frac{\text{تغاير (س ، ص)}}{\text{تباين س}} \quad (8)$$

وفى إيجاد ب من (٥) أو (٨) يمكن أن نجمع أو نطرح أى عدد من جميع قيم س وأى عدد من جميع قيم ص دون أن تتأثر قيمة ب . وفى حالة التوزيعات التكرارية حيث يكون لكل زوج ($\text{س}_{\text{ر}}$ ، $\text{ص}_{\text{ر}}$) تكرار $\text{ك}_{\text{ر}}$ نحصل على ب من أى من الصيغتين بوضع $\text{ك}_{\text{ر}}$ بعد كل رمز مج .

مثال (٩ – ١) :

أوجد معادلة انحدار ص على س من البيانات الآتية :

س :	١,٥	١,٨	٢,٤	٣,٠	٣,٥	٣,٩	٤,٤	٤,٨	٥,٠
ص :	٤,٨	٥,٧	٧,٠	٨,٣	١٠,٩	١٢,٤	١٣,١	١٣,٦	١٥,٣

ومنها أوجد أحسن تقدير لقيمة ص عندما س = ٣,٧ .

الحل :

(يلاحظ أننا لو رسمنا شكل الانتشار لوجدنا أن افتراض الخطية هو افتراض معقول) . يتطلب الحل إيجاد كل من مج س ، مج ص ، مج س^{2} ، مج س ص وهذه جميعاً نحصل عليها من الجدول (٩ – ١) الآتى .

لاحظ أن العمود الأخير لا لزوم له في إيجاد معادلة انحدار ص على س ولكنه سيلزمنا فيما بعد في تحليل الانحدار . ويمكن استخدام حاسبات الجيب للحصول على هذه المجاميع دون ضرورة لكتابة التفاصيل المبينة بالجدول وذلك توفيرا للوقت والجهد .

جدول (٩ – ١)

إيجاد معادلة خط الانحدار من بيانات المثال (٩ – ١)

س	ص	س ص	$س^2$	$ص^2$
١,٥	٤,٨	٧,٢٠	٢,٢٥	٢٣,٠٤
١,٨	٥,٧	١٠,٢٦	٣,٢٤	٣٢,٤٩
٢,٤	٧,٠	١٦,٨٠	٥,٧٦	٤٩,٠٠
٣,٠	٨,٣	٢٤,٩٠	٩,٠٠	٦٨,٨٩
٣,٥	١٠,٩	٣٨,١٥	١٢,٢٥	١١٨,٨١
٣,٩	١٢,٤	٤٨,٣٦	١٥,٢١	١٥٣,٧٦
٤,٤	١٣,١	٥٧,٦٤	١٩,٣٦	١٧١,٦١
٤,٨	١٣,٦	٦٥,٢٨	٢٣,٠٤	١٨٤,٩٦
٥,٠	١٥,٣	٧٦,٥٠	٢٥,٠٠	٢٣٤,٠٩
٣٠,٣	٩١,١	٣٤٥,٠٩	١١٥,١١	١٠٣٦,٦٥

من الصيغة (٥) نجد أن ب = $\frac{٩ \times ٣٤٥,٠٩ - ٣٠,٣ \times ٩١,١}{٩ \times ١١٥,١١ - (٣٠,٣)^2}$ = ٢,٩٣٠٣

كما أن $\bar{س} = \frac{٣٠,٣}{٩}$ = ٣,٣٦٦٧ ، $\bar{ص} = \frac{٩١,١}{٩}$ = ١٠,١٢٢٢

∴ معادلة انحدار ص على س (من الصيغة ٧) هى

$\hat{ص}$ = ١٠,١٢٢٢ + ٢,٩٣٠٣ (س – ٣,٣٦٦٧)

ومنها $\hat{ص} = ٠,٢٥٦٨ + ٢,٩٣٠٣$ س

وحينما س = ٣,٧ فإن أحسن تقدير لقيمة ص هي :

$\hat{ص} = ٠,٢٥٦٨ + ٢,٩٣٠٣ \times ٣,٧ = ١١,٠٩٨٩$

(٩ – ٤) الخطأ المعياري للتقدير Standard Error of Estimate

كما هو الحال عند دراسة بيانات عن متغير واحد حيث نصف هذه البيانات بواسطة الوسط الحسابي الذي يعطي تقديراً لمتوسط هذه البيانات ثم ندعم هذا الوصف بتقدير التشتت بواسطة الانحراف المعياري ، فاننا نصف البيانات ذوات المتغيرين بواسطة خط الانحدار الذي يعطي تقديراً لمتوسطات قيم ص عند قيم س ونستكمل هذا الوصف بتقدير مدى تشتت نقط شكل الانتشار حول هذا الخط . ويقاس هذا التشتت بما يسمى بالخطأ المعياري لتقدير ص من معادلة الانحدار (٧) ، ويرمز له بالرمز $ع_{ص.س}$ ويعرف كالآتي :

$$ع^{٢}_{ص.س} = \frac{١}{ن - ٢} \sum (ص_ر - \hat{ص}_ر)^{٢} \qquad (٩)$$

وهذا المقياس له أهمية كبرى في عمليات الاستنتاج الإحصائي كما سنرى بعد .

ملاحظة :

يمكن إثبات أن .

$$ع^{٢}_{ص.س} = \frac{١}{ن - ٢} (\sum ص^{٢} - أ \sum ص - ب \sum س ص) \qquad (١٠)$$

وهذه الصيغة أسهل في حساب الخطأ المعياري من الصيغة (٩) وتستخرج من نفس المجاميع التي نوجدها لحساب معادلة الانحدار .

مثال (٩ – ٢) :

أوجد الخطأ المعيارى لخط الانحدار الناتج في المثال (٩ – ١) السابق .

الحل :

نعوض في الصيغة (١٠) من مجاميع الجدول (٩ – ١) بالقيمتين
أ = ٠,٢٥٦٨ ، ب = ٢,٩٣٠٢

$$ع^2_{ص.س} = \frac{1}{7}(١٠٣٦,٥٦ - ٠,٢٥٦٨ \times ٩١,١ - ٢,٩٣٠٣ \times ٣٤٥,٠٩)$$

$$= ٠,٢٩١٢$$

$$\therefore ع_{ص.س} = ٠,٥٣٩٦$$ (الخطأ المعيارى مقدراً من العينة)

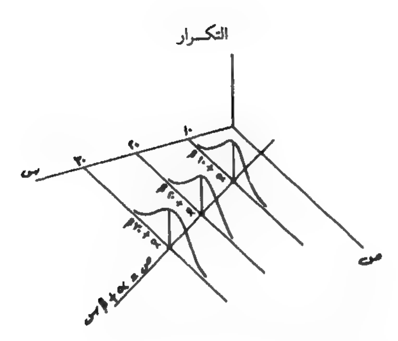
(٩ – ٥) استنتاجات إحصائية :

إن إيجاد معادلة الانحدار الخطى لم يستلزم إلا وضع الافتراضين الأول والثاني السابق ذكرهما ، أما إذا أردنا القيام باستنتاجات إحصائية أى إصدار قرارات عن المجتمع عن طريق العينة فإن ذلك يتطلب وضع افتراض خاص عن توزيع احتمال المتغير العشوائي ص . وفى المعتاد نضع الافتراض الآتي (الذى يمكن هو أيضاً اختبار صحته) .

الافتراض الثالث :

« عند أى قيمة ثابتة س يكون للمتغير العشوائي ص توزيع معتدل متوسطه $\alpha + \beta$ س وتباينه عدد ثابت مجهول σ^2 مستقل عن س » .

ويمكن تصوير هذا الافتراض كالآتى :



الشكل (٩ – ٤) توزيعات معتدلة للمتغير ص عند س = ١٠ ، ٢٠ ، ٣٠

ويلاحظ ما يلي :

(أ) متوسطات التوزيعات الصادية تختلف باختلاف قيم س وهي تقع جميعها على خط الانحدار .

(ب) جميع التوزيعات الصادية متوازية ولها نفس الشكل لأن لها نفس التباين وتختلف فقط في مواقعها كما جاء في (أ) .

من الافتراضات الثلاثة المذكورة ، بالاضافة إلى افتراض العشوائية الذي يضمن استقلال وحدات التجريب عن بعضها البعض ، يمكن أن نخرج بعدة استنتاجات نختار منها ما يلي :

(أولا) اختبار الفرض β = ك حيث ك عدد معين .

للوصول إلى خط الانحدار صَ = أ + ب س نوجد معامل الانحدار ب من واقع بيانات مأخوذة من عينة ما ، وإذا اخترنا عينة أخرى نحصل على قيمة مختلفة

لهذا المعامل . أى أن قيمة ب تختلف من عينة إلى أخرى ، ولذلك نعتبر أن أى قيمة للمعامل ب هى إحدى القيم المشاهدة من متغير عشوائى B ذى عدد غير منتهى من القيم . وتحت الافتراضات الثلاثة للانحدار يمكن إثبات أن لهذا المتغير توزيعا معتدلا وسطه الحسابى β وانحرافه المعيارى يقدر بالمقدار $ع_{ص.س} / \sqrt{م^2_س}$ حيث $ع_{ص.س}$ هو الخطأ المعيارى للتقدير المعطى بالصيغة (٩) وحيث

$$م^2_س = مج (س - \bar{س})^2 = مج س^2 - \frac{(مج س)^2}{ن}$$

وينتج من البند (٦ — ٦) أن الإحصاءة

$$ت = \frac{B - \beta}{ع_{ص.س} / \sqrt{م^2_س}} \qquad (11)$$

يكون لها توزيع ت بدرجات حرية (ن — ٢) . وبهذه الإحصاءة نستطيع كالمعتاد اختبار الفرض الصفرى ف. : β = ك ضد أى فرض آخر وذلك بإيجاد ت. من بيانات العينة أى بوضع B = ب وعلى أساس صحة الفرض الصفرى أى بوضع β = ك ثم مقارنة هذه القيمة بالقيمة الحرجة $ت_{\alpha[ن-2]}$ التى نستخرجها من جدول ت .

نتيجة : اختبار خطية العلاقة بين س ، ص .

بصفة خاصة نستطيع اختبار الفرض الصفرى ف. : β = ٠ ضد الفرض $\beta \neq$. فإذا أشار الاختبار بقبول الفرض الصفرى حكمنا بعدم وجود علاقة خطية بين س ، ص لأن معنى β = ٠ أن تكون ص = α أى ص تكون ثابتة ولا يكون لقيم س أى أثر خطى على ص . أما إذا رفضنا الفرض الصفرى لصالح الفرض

الآخر $\beta \neq$ ٠ نحكم بأن هناك علاقة خطية بين المتغيرين وإن كان هذا لا يمنع من وجود علاقة أخرى قد تكون أفضل من العلاقة الخطية مثل ص $= \alpha +$ β س + γ س^{٢}. وهناك اختبار آخر نعرف منه مدى انحراف العلاقة الحقيقية بين س ، ص عن العلاقة الخطية ، وسنقدم هذا الاختبار فى البند (٩ – ٩) .

مثال (٩ – ٣) :

ابحث وجود علاقة خطية بين المتغيرين س ، ص من بيانات المثال (٩ – ١) مستخدماً مستوى الدلالة ٠,٠١

الحل :

الفرض الصفرى ف$_{٠}$: β = ٠ (لا يوجد علاقة خطية)

الفرض الآخر ف$_{١}$: $\beta \neq$ ٠ (اختبار ذو جانبين)

$$م^{٢}_{س} = مج\, س^{٢}_{ر} - \frac{(مج\, س_{ر})^{٢}}{ن} = ١١٥,١١ - \frac{(٣٠,٣)^{٢}}{٩}$$

$$= ١٣,١$$

$$\therefore \sqrt{م^{٢}_{س}} = ٣,٦١٩٤$$

كما أن ع$_{ص|س}$ = ٠,٥٣٩٦ وقد سبق إيجادها

من (١١) وعلى أساس لأن $\beta_{٠}$ = ٠ نجد أن :

$$ت_{ى} = \frac{٣,٦١٩٤ \times ٢,٩٣٠٣}{٠,٥٣٩٦} = ١٩,٦٥٥٢$$

من الجدول ت$_{٠,٠١ \,(٧)}$ = ٣,٤٩٩

نرفض ف$_{٠}$ ونحكم بأن هناك علاقة خطية بين المتغيرين .

مثال (٩ – ٤) :

في بيانات المثال (٩ – ١) ابحث ما إذا كانت β < ٢,٥ عند مستوى الدلالة ٠,٠١

الحل :

ف. : β = ٢,٥

ف١ : β < ٢,٥   (اختبار ذو جانب واحد)

من (١١) وعلى أساس صحة الفرض الصفرى β = ٢,٥ نجد أن

$$ت_ى = \frac{(٢,٩٣٠٣ - ٢,٥) \times ٣,٦١٩٤}{٠,٥٣٩٦} = ٢,٨٨٦٣$$

من الجدول : $ت_{٠,٠٢ \; [٧]} = ٢,٩٩٨$

نقبل الفرض الصفرى أن β = ٢,٥ عند المستوى ٠,٠١ ونحكم بأن β لا تزيد عن ٢,٥ .

(ثانياً) فترات الثقة للبارامتر β .

من الإحصاءة (١١) نستطيع أن نثبت أن العددين

$$ب \pm \frac{ع_{ص.س}}{\sqrt{م_س م}} \times ت_{\alpha \; [ن - ٢]} \qquad (١٢)$$

هما حدا الثقة بدرجة (α – ١) للبارامتر β .

(ثالثاً) فترات الثقة للقيمة الحقيقية للمتغير ص عند قيمة معينة $س_ر$:

يمكن إثبات أن العددين

$$\hat{ص} \pm ع_{ص.س} \sqrt{1 + \frac{1}{ن} + \frac{(س_ر - \bar{س})^2}{م م_س}} \times ت_{\alpha[ن-2]} \qquad (13)$$

هما حدا الثقة بدرجة $(1 - \alpha)$ لقيمة ص عندما تأخذ س القيمة $س_ر$.

مثال (9 – 5) :

للمثال (9 – 1) أوجد فترة ثقة بدرجة 95٪ لكل من :

(أ) البارامتر β (ب) القيمة الحقيقية للمتغير ص عند س = 2

الحل :

(أ) نعوض في (12) مع ملاحظة أن $ت_{0.05[7]} = 2.365$

$$\text{الحد الأدنى للفترة} = 2.9303 - \frac{0.0396}{3.6194} \times 2.365 = 2.578$$

$$\text{الحد الأعلى للفترة} = 2.9303 + \frac{0.0396}{3.6194} \times 2.365 = 3.283$$

إذن الفترة (2.578 ، 3.283) هى فترة ثقة بدرجة 95٪ للبارامتر β .

(ب) نعوض في (13) ، مع ملاحظة أنه عندما س = 2 فإن

$$\hat{ص} = 0.2568 + 2.9303 \times 2 = 6.1174$$

$$\text{الحد الأدنى للفترة} = 6.1174 - 0.0396 \sqrt{1 + \frac{1}{9} + \frac{(2 - 3.3667)^2}{13.1}} \times 2.365$$

$$= 6.1174 - 0.0396 \times 1.1197 \times 2.365 = 4.689$$

الحد الأعلى للفترة = ٧,٥٤٦

إذن الفترة (٤,٦٨٩ ، ٧,٥٤٦) هي فترة ثقة بدرجة ٩٥٪ لقيمة ص عند س = ٢

ملاحظة :

هناك برامج جاهزة للحاسب الالكتروني تعطى جميع القيم المطلوبة في تحليل الانحدار بدءا من قيم الصيغة (٥) إلى قيم الصيغة (١٣) بمجرد تغذيته بالبيانات الخام .

(٩ – ٦) التوسع في استخدام الانحدار الخطى البسيط :

يمكن تناول بعض العلاقات غير الخطية بنفس الطريقة التي نتناول بها العلاقات الخطية وذلك باختيار تحويلات مناسبة للمتغيرات بحيث تأخذ العلاقة المعطاة الصورة الخطية . وكمثال لذلك نفرض أن العلاقة بين المتغيرين س ، ص على الصورة :

$$ص = \alpha \, هـ^{\beta \, س} \quad حيث \ \alpha > ٠ \ ، \ ص > ٠$$

ونريد تقدير α ، β من العينة . بأخذ لوغاريتم كل من الطرفين للأساس هـ نجد أن :

$$لو \ ص = لو \ \alpha + \beta \ س$$

$$أو \ \bar{ص} = \bar{\alpha} + \beta \ س$$

$$حيث \ \bar{\alpha} = لو \ \alpha \ ، \ \bar{ص} = لو \ ص$$

وهذه العلاقة الأخيرة على صورة خطية . لإيجاد تقديرين ا ، ب للمجهولين α ، β نبدأ بتحويل كل قيمة ص إلى لو ص فيصبح لدينا الأزواج (س ، صَ) بدلا من الأزواج (س ، ص) . نوجد التقديرين ب ، اَ كالمعتاد من الصيغتين (٥ ، ٦) فتكون ب تقديراً للبارامتر β وتكون ا = هـ^اَ هى تقدير للبارامتر α وتكون معادلة الانحدار هى :

$$\text{ص} = \text{ا هـ}^{\text{ب س}}$$

وبنفس الطريقة يمكن تناول المعادلات الآتية :

ص = α + $\frac{\beta}{\text{س}}$ حيث س $\neq$ ٠ (ضع سَ = $\frac{1}{\text{س}}$ التحويل القالب)

ص = α س$^{\beta}$ حيث س > ٠ ، α > ٠ ، ص > ٠ (تحويل لوغاريتمي)

ص = α + β حتا ٣ س (ضع سَ = حتا ٣ س)

ص س$^{\alpha}$ = β حيث س > ٠ ، ص > ٠ ، β > ٠ (تحويل لوغاريتمي)

مثال (٩ – ٦) :

المعروف أن العلاقة بين ضغط الغاز ض وحجمه ح تأخذ الصورة ض ح$^{\beta}$ = α أوجد تقديراً للبارامترين المجهولين α ، β من البيانات التجريبية الآتية ثم أوجد أحسن تقدير لضغط الغاز حين يكون حجمه ح = ١٠٠

ح :	١٩٤,٠	١١٨,٦	٨٨,٧	٧٢,٤	٦١,٨	٥٤,٣
ض :	١٠,١	١٩,٢	٢٨,٤	٣٧,٦	٤٩,٥	٦١,٢

الحل :

نحول المعادلة ص ح$^{\beta}$ = α إلى صورة خطية بأخذ لوغاريتم كل من الطرفين للأساس ١٠ .

∴ لو ض + β لو ح = لو α أو لو ض = لو α − β لو ح

أو ص = $\acute{\alpha}$ + $\acute{\beta}$ س

حيث ص = لو ض ، $\acute{\alpha}$ = لو α ، $\acute{\beta}$ = −β ، س = لو ح

لتقدير بارامترات هذه المعادلة الخطية ينبغي أن نوجد لوغاريتم كل قيمة معطاة من قيم الحجم ح ولوغاريتم كل قيمة مناظرة من قيم الضغط ص . وباستخدام جداول اللوغاريتمات أو الحاسبات نحصل على العمودين الأول والثاني من الجدول الآتي ، ثم نستكمل الحساب كما في المثال (٩ − ١) .

الجدول (٩ − ٢)

س = لو ح	ص = لو ض	س^{2}	س ص
٢,٢٨٧٨	١,٠٠٤٣	٥,٢٣٤٠	٢,٢٩٧٦
٢,٠٧٤١	١,٢٨٣٣	٤,٣٠١٩	٢,٦٦١٧
١,٩٤٧٩	١,٤٥٣٣	٣,٧٩٤٣	٢,٨٣٠٩
١,٨٥٩٧	١,٥٧٥٢	٣,٤٥٨٥	٢,٩٢٩٤
١,٧٩١٠	١,٦٩٤٦	٣,٢٠٧٧	٣,٠٣٥٠
١,٧٣٤٨	١,٧٨٦٨	٣,٠٠٩٥	٣,٠٩٩٧
١١,٦٩٥٣	٨,٧٩٧٥	٢٣,٠٠٥٩	١٦,٨٥٤٣

$\bar{س} = ١,٩٤٩٢$ ، $\bar{ص} = ١,٤٦٦٣$

على فرض أن $\bar{أ}$ ، $\bar{ب}$ هما التقديران المطلوبان للثابتتين α ، β نجد ما يلى :

من الصيغة (٥) : $\bar{ب} = \frac{٦ \times ١٦,٨٥٤٣ - ١١,٦٩٥٣ \times ٨,٧٩٧٥}{٦ \times ٢٣,٠٠٥٩ - (١١,٦٩٥٣)^{٢}}$

$= -١,٤٠$

من الصيغة (٦) : $\bar{أ} = \bar{ص} - \bar{ب}\,\bar{س} = ١,٤٦٦٣ + ١,٤٠ \times ١,٩٤٩٢$

$= ٤,١٩٥٢ = ٤,٢$ تقريبا

إذن $\qquad ب = -\bar{ب} = ١,٤٠$

، $\qquad أ = ١٠^{٤,٢} = ١٥٨٤٨,٩ = ١٥٨٤٩$ تقريباً

وهذان هما التقديرات المطلوبان للثابتين β ، α وعلى ذلك فإن المعادلة التي تربط الحجم والضغط الناتجة من العينة هي :

$$ص\ ح^{١,٤٠} = ١٥٨٤٩$$

وحين ح = ١٠٠ فإن أحسن تقدير للضغط ض ينتج كالآتى :

ض $\times ١٠٠^{١,٤٠} = ١٥٨٤٩$ ومنها ض $\times ١٠^{٢,٨٠} = ١٥٨٤٩$

إذن ض $\qquad = ٢٥,١٢$ تقريباً .

تمارين (٩ – ١)

من كل من البيانات المبينة في المسائل الخمسة الآتية :

(ا) أوجد معادلة انحدار ص على س .

(ب) أوجد الخطأ المعياری $ع_{ص.س}$ لخط الانحدار .

(حـ) ابحث ما إذا كان هناك علاقة خطية بين المتغيرين س ، ص .

(ى) إذا ثبت وجود علاقة خطية فأوجد فترة ثقة بدرجة ٩٥٪ للقيمة الحقيقية للمتغير ص عند القيمة $س_{.}$ المعطاة .

(١)

س :	٢,٨	٢,٩	٣,٠	٣,١	٣,٢	٣,٢	٣,٢	٣,٣	٣,٤
ص :	٢٧	٢٣	٣٠	٢٨	٣٠	٣٢	٣٤	٣٣	٣٠

حيث س ترمز إلى كثافة الحديد الخام (جرام / $سم^{٣}$)

، ص ترمز إلى النسبة المئوية لمحتوى الحديد .

، $س_{.}$ = ٣

(٢)

س :	٤٨	٤٩	٥٠	٥١	٥٢
ص :	٢٧٧,١	٢٧٩,٣	٢٨١,٤	٢٨٣,٢	٢٨٤,٨

حيث س ترمز إلى طول الطفل عند الولادة بالسنتيمترات

، ص ترمز إلى مدة الحمل بالأيام .

، $س_{.}$ = ٥٠

(٣)

س :	٨	٩	٩	١٢	١٣	١٤	١٤	١٦	١٦	١٦
ص :	-٥٥	-٥٥	-٥٣	-٥٠	-٥٠	-٣٥	-٣٧	-٣٥	-٢٧	-٢٤

حيث س ترمز إلى درجات الحرارة بالسنتيجراد

، ص ترمز إلى الانحراف (بمضاعفات الـ $\frac{١}{١٠٠٠}$ زاوية دائرية) لنوع معين من الرؤية التلسكوبية .

، $\bar{س}$ = ١٣,٥

(٤) س :	١٥	٣٠	٦٠	١٢٠	٣٠٠	٦٠٠
ص :	٤,٠	٦,٣	٥,٢	٦,٤	٥,٥	٦,٤

حيث س ترمز إلى المدة بالثواني

، ص ترمز إلى مقدار المادة التي تبددت (بالجرامات في اللتر) من استحلاب الزئبق في محلول زيتريت الصديون بتأثير الاهتزاز الناتج من الصوت عن ذبذبات فوق الصوتية في المدة س .

، $\bar{س}$ = ٧٠

(٥) س :	٠,٠١	٠,٠٢	٠,٠٤	٠,٠٨	٠,١٥	٠,٣٠	٠,٥٠	١,٠٠	٢,٠٠
ص :	١٢	١٨	٣	٢٧	١٦	١٣	١٠	٧٠	٩٠

حيث س ترمز إلى النسبة المئوية لدرجة تركيز الكلورونافتالين

، ص ترمز إلى النسبة المئوية لموت النمل الأبيض .

، $\bar{س}$ = ٠,٢٥

(٦) في عينة عشوائية من ٨ أزواج من القيم ($س_ر$ ، $ص_ر$) وجد أن معامل الانحدار

ب = ٠,٣١ وأن $ع^{٢}_{ص.س}$ = ٠,٤ ، $\sqrt{ع^{٢}_{س}}$ = ٠,٣

اختبر الفرض أن β = ٠ ضد الفرض أن $\beta >$.

(٩ – ٧) معنى آخر للانحدار – تحليل التباين :

فى البنود السابقة من هذا الفصل كنا نأخذ الانحدار على أنه وسيلة لإيجاد معادلة تمكننا من معرفة أحسن تقدير لقيمة متغير عشوائى ص عن طريق قيمة معطاة لمتغير س . ولذلك وصفنا معادلة الانحدار بأنها معادلة تنبؤ . إلا أن الانحدار يؤخذ أيضا على أنه وسيلة لتفسير الاختلاف المشاهد فى قيم المتغير ص ، وذلك استجابة لتساؤل هام عن العوامل المؤثرة فى هذا الاختلاف ، وبصفة خاصة عما إذا كان هذا الاختلاف يرجع إلى التغير فى قيم المتغير س أم يرجع إلى عوامل عشوائية أو عوامل أخرى لا تتعلق بالمتغير س .

إن الإجابة عن هذا التساؤل تتطلب تحليل الاختلاف فى قيم ص وهو ٢ ٢ (ص) $= \sum (ص_ر - \bar{ص})^2$ إلى مركبتين مستقلتين تعتمد أحدهما على قيم المتغير س ولا تعتمد الأخرى عليها ثم تقييم كل من هاتين المركبتين .

ولبيان كيفية هذا التحليل نعتبر المتساوية الآتية :

$$ص_ر - \bar{ص} = ص_ر - \bar{ص} + \hat{ص}_ر - \hat{ص}_ر$$

$$= (\hat{ص}_ر - \bar{ص}) + (ص_ر - \hat{ص}_ر) \quad \text{حيث } ر = ١ ، ٢ ، \ldots ، ن$$

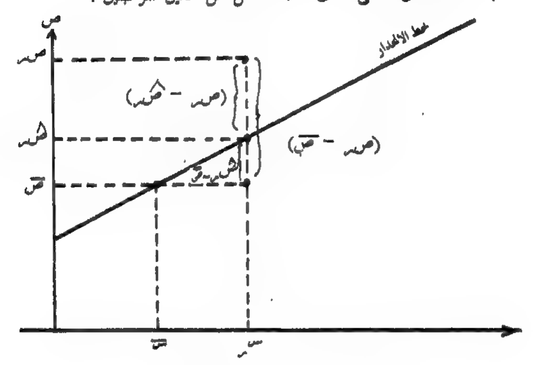
إن هذه المتساوية واضحة جبريا ، كما أنها تتضح هندسيا من الشكل (٩ – ٥) مع ملاحظة أن النقطة $(\bar{س} ، \bar{ص})$ تقع دائما على خط الانحدار . وبتربيع طرفى المتساوية ثم الجمع تنتج المتساوية الآتية :

$$\sum_{ر=١}^{ن} (ص_ر - \bar{ص})^2 = \sum_{ر=١}^{ن} (\hat{ص}_ر - \bar{ص})^2 + \sum_{ر=١}^{ن} (ص_ر - \hat{ص}_ر)^2 \quad (١٤)$$

مع ملاحظة أن الحد الأوسط فى عملية تربيع الطرف الأيسر ينعدم عند عملية

الجمع . وبذلك نكون قد جزأنا الاختلاف الكلى فى ص إلى مركبتين بطريقة مشابهة لتقسيم الاختلاف الكلى فى تحليل التباين .

لنبحث الآن فى المعنى الذى تتضمنه كل من هاتين المركبتين .



الشكل (٩ – ٥) : تجزيء الاختلاف فى المتغير ص

الاختلاف المفسر والاختلاف غير المفسر

EXPLAINED AND UNEXPLAINED VARIATION

من الشكل (٩ – ٥) نرى أننا إذا رسمنا خط الانحدار وحددنا قيمة معينة سر فإن قيمة ص̂ر المناظرة تكون قد تحددت تماما لأنها تقع على خط الانحدار ويكون الفرق (ص̂ر − ص̄) بين ص̂ر والقيمة الثابتة ص̄ هو فرق يعتمد كلية على قيمة سر وبالتالى فإن مجموع مربعات الفروق مج (ص̂ر − ص̄)٢ يعتمد كلية على قيم المتغير س . ولما كان هذا المقدار هو جزء من الاختلاف الكلى فى ص كما يظهر من المتساوية (١٤) فإن مج (ص̂ر − ص̄)٢ يكون هو الجزء من الاختلاف فى ص

الذى يُعزّى إلى التغير فى س ، أى إلى التغير الذى حدث فى ص نتيجة للتغير فى س ، وهـذا ما نسميـة بانحـدار ص على س ولذلك نرمز لهـذا الاختلاف بالرمـز م م (الانحـدار الخطى) وله درجة واحدة من درجات الحريـة . ولما كـان هـذا الاختلاف مصدره معروف (وهو التغير فى س) فقد اصطلح على تسميته أيضا بالاختلاف المفسر ونكتبه رمزيا كالآتى :

الاختلاف المفسر = م م (الانحدار) = مجـ (صَ^ر – صَ)٢ بدرجة حرية واحدة (١٥)

ويمكن إثبات أن هذا الاختلاف يمكن أن يحسب كالآتى :

الاختلاف المفسر = ب٢ م م (س) حيث ب معامل الانحدار (١٦)

أو = [م حـ (س ، ص)]٢ / م م (س) (١٧)

حيث م م (س) = مجـ (س – سَ)٢ = مجـ س٢ – (مجـ س)٢ / ن

، م حـ (س ، ص) = مجـ (س – سَ)(ص – صَ) = مجـ س ص – (مجـ س)(مجـ ص) / ن

أما المركبة الثانية وهى مجـ (صر – صَ^ر)٢ فهى الاختلاف الذى يتبقى بعد طرح الاختلاف المفسر من الاختلاف فى ص ، وهو يعتمد على عوامل مجهولة لا تتعلق بالمتغير س . وقد اصطلح على تسمية هذا الاختلاف بالاختلاف المتبقى residual variation أو بالاختلاف غير المفسر . ومع تذكر أن صر هى قيم عشوائية فإن هذا الاختلاف يعبر عن التشتت غير المنتظم لنقط شكل الانتشار حول خط الانحدار أى عن الانحرافات الرأسية حول خط الانحدار وسنرمز له بالرمز م م (الانحراف عن خط الانحدار) وله ن – ٢ من درجات الحرية . أى أن

الاختلاف غير المفسر = م م (الانحراف عن خط الانحدار)

$$= مج (ص_ر - \hat{ص}_ر)^٢$$

$$= م م (ص) - الاختلاف المفسر \qquad (١٨)$$

والمعنى الذى يتضمنه هذا الاختلاف يبرر استخدامه كأساس لحساب الخطأ المعيارى $ع_{ص.س}$ الوارد فى الصيغة (٩) بالبند (٩ - ٤) ، حيث

$$ع^٢_{ص.س} = \frac{مج (ص_ر - \hat{ص}_ر)^٢}{ن - ٢} = \frac{الاختلاف غير المفسر}{درجات الحرية} \qquad (١٩)$$

نستطيع حينئذ أن نكتب المتساوية (١٤) كالآتى :

م م (ص) = م م (الانحدار الخطى) + م م (الانحراف عن خط الانحدار)(٢٠)
بدرجات حرية ن - ١ ، ١ ، ن - ٢ على الترتيب .

ويفضل تسجيل قيم هذه المتساوية فى جدول التباين الآتى .

الجدول (٩ - ٣)

جدول التباين للانحدار

مصدر التباين	م م	د ح	ط م	ف
الاختلاف المفسر (الانحدار الخطى)	$مج (\hat{ص}_ر - \bar{ص})^٢$	١	$ع^٢_١$	$\frac{ع^٢_١}{ع^٢_٢}$
الاختلاف غير المفسر (الانحراف عن الخطية)	$مج(ص_ر - \hat{ص}_ر)^٢$	ن - ٢	$ع^٢_٢$	
الاختلاف الكلى فى ص	$مج(ص_ر - \bar{ص})^٢$	ن - ١		

بعد تجزىء الاختلاف الكلى فى ص بهذه الصورة يبقى أن نختبر ما إذا كان الانحدار الخطى قد فسر جزءا ذا بال من هذا الاختلاف ، أى أن نختبر ما إذا كان التباين المفسر أكبر كبرا جوهريا من تباين الاختلاف غير المفسر . وهذا ما نقيسه باختبار ف بالصيغة الآتية بشرط توفرافتراضاتالانحدار ، ومع ملاحظة أن هذين التباينين هما تقديران مستقلان للتباين σ^2 لتوزيع المتغير ص .

$$ف = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}} \quad \text{بدرجتى حرية ١، ن − ٢} \quad (٢١)$$

وهذا الاختبار يكافىء اختبار ت الوارد بالصيغة (١١) بالبند (٩ − ٥) لاختبار وجود علاقة خطية بين المتغيرين س ، ص أى لاختبار الفرض الصفرى $\beta = ٠$ ، ضد الفرض $\beta \neq ٠$. ويمكن اشتقاق أى منهما من الآخر ، إذ أن قيمة ف بالصيغة (٢١) تساوى مربع قيمة ت بالصيغة (١١) . وبذلك نكون قد توصلنا إلى طريقة أخرى لاختبار وجود علاقة خطية بين المتغيرين .

وجدير بالذكر أن الصيغة (٢١) هى صيغة عامة لاختبار مدى دقة التنبؤ من خط الانحدار مهما كان عدد المتغيرات المستقلة المستخدمة فى عملية التنبؤ، وتختلف طريقة حساب البسط والمقام فى هذه الصيغة باختلاف عدد المتغيرات التنبؤية.

مثال (٩ − ٧) :

ابحث وجود علاقة خطية بين المتغيرين س ، ص من بيانات المثال (٩ − ١) مستخدما طريقة تحليل التباين .

الحل :

الفرض الصفرى : $\beta = ٠$ ، الفرض الآخر : $\beta \neq ٠$.

من الجدول (٩ – ١) نستطيع حساب القيم اللازمة لاستخدام الصيغة (٢١) كالآتى :

$$\text{م}^2\ (\text{ص}) = ١٠٣٦,٦٥ - \frac{٩١,١^2}{٩} = ١١٤,٥١٥٦$$

$$\text{م}^2\ (\text{س}) = ١١٥,١١ - \frac{(٣٠,٣)^2}{٩} = ١٣,١$$

$$\text{م ص}\ (\text{س ، ص}) = ٣٤٥,٠٩ - \frac{٩١,١ \times ٣٠,٣}{٩} = ٣٨,٣٨٦٧$$

من (١٧) : الاختلاف المفسر $= \frac{(٣٨,٣٨٦٧)^2}{١٣,١} = ١١٢,٤٨٣٩$ $\quad \nu_1 = ١$

∴ الاختلاف غير المفسر = ١١٤,٥١٥٦ – ١١٢,٤٨٣٩

= ٢,٠٣١٧ $\quad \nu_2 = ٧$

وينشأ جدول التباين الآتى :

الجدول (٩ – ٤)

مصدر التباين	م ٢	د ح	ط ٢	ف ى
الاختلاف المفسر	١١٢,٤٨٣٩	١	١١٢,٤٨٣٩	٣٨٧,٦٠٨**
الاختلاف غير المفسر	٢,٠٣١٧	٧	٠,٢٩٠٢	
الاختلاف الكلى	١١٤,٥١٥٦	٨		

إن القيمة $ف_ى$ = ٣٨٧,٦٠٨ أكبر بكثير من القيمة الحرجة $ف_{٠,٠١[١،٧]}$ = ١٢,٢ مما يجعلنا نرفض الفرض الصفرى عند مستوى عالى من الدلالة ونحكم بوجود علاقة خطية بين المتغيرين .

نلاحظ أن الجذر التربيعى لقيمة $ف_ى$ وهو $\sqrt{٣٨٧,٦٠٨}$ = ١٩,٦٨٨ يساوى قيمة $ت_ى$ = ١٩,٦٥٥٢ السابق إيجادها بالمثال (٩ – ٣) ويرجع الفرق بين القيمتين إلى أخطاء التقريب .

معامل التحديد **COEFFICIENT OF DETERMINATION**

استكمالا للانحدار كوسيلة لتفسير الاختلاف فى قيم المتغير التابع ص يهمنا أن نقدر نسبة الاختلاف الذى فسره الانحدار الخطى إلى الاختلاف الكلى فى المتغير ص . وهذه النسبة تسمى بمعامل التحديد ويرمز لها بالرمز $ر^٢$. أى أن :

$$ر^٢ = \frac{\text{الاختلاف المفسر}}{\text{الاختلاف الكلى}} \qquad (٢٢)$$

ففى المثال السابق نجد أن :

$$ر^٢ = \frac{١١٢,٤٨٣٩}{١١٤,٥١٥٦} = ٠,٩٨٢ \text{ تقريبا}$$

وهذا يعنى أن الانحدار الخطى قد فسر حوالى ٩٨,٢٪ من الاختلاف الكلى أى جزءا جوهريا منه ، مما يؤكد صحة العلاقة الخطية بين المتغيرين .
ويرمز لمعامل التحديد بالرمز $ر^٢$ لأنه يساوى مربع معامل الارتباط ر الذى سنتناوله فى الفصل التالى . ويلاحظ أن

$$١ \geqslant ر^٢ \geqslant ٠$$

لأن كلا من البسط والمقام فى التعريف (٢٢) هو مجموع مربعات لا يمكن أن يكون سالبا وإذن $ر^٢ \geqslant ٠$ كما أن البسط جزء من المقام كما يتضح من المتساوية

(١٤) وإذن ر٢ ≥ ١ . وإذا كانت ر٢ = ٠ فإن هذا يعنى أن الانحدار الخطى لا يفسر شيئا من الاختلاف فى ص ولا يكون للمتغير فى س أى أثر فى التغير فى ص . أما إذا كانت ر٢ = ١ فإن هذا يعنى أن الانحدار قد فسر الاختلاف فى ص بأكمله ، وهذا يحدث حين تكون القيم المشاهدة صر مساوية للقيم المناظرة صَر المقدرة من معادلة الانحدار ، أى تكون النقط (سر ، صر) المشاهدة واقعة جميعها على خط الانحدار . وحين تكون قيمة ر٢ صغيرة فإن هذا يعنى أن الجزء الأكبر من الاختلاف فى صّ يرجع إلى عوامل ومتغيرات لا تدخل فى الانحدار أى لا علاقة لها بالتغير فى المتغير المستقل س .

من الصيغة (١٧) يمكن أن نكتب معامل التحديد بدلالة مجاميع المربعات ومجاميع حواصل الضرب كالآتى :

$$ر^{٢} = \frac{[\text{م ح (س ، ص)}]^{٢}}{\text{م م (س) . م م (ص)}} \qquad (٢٣)$$

هذا ويمكن كتابة اختبار ف بالصيغة (٢١) بدلالة معامل التحديد كالآتى :

$$ف = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}} = \frac{\text{الاختلاف المفسر} \div ١}{\text{الاختلاف غير المفسر} \div (ن - ٢)}$$

$$= \frac{\text{الاختلاف المفسر}}{(\text{الاختلاف الكلى} - \text{الاختلاف المفسر}) \div (ن - ٢)}$$

وبقسمة كل من البسط والمقام على الاختلاف الكلى م م (ص) ينتج أن

$$ف = \frac{ر^{٢}}{(١ - ر^{٢})/(ن - ٢)} \qquad (٢٤)$$

وهذه صيغة ثالثة لاختبار وجود علاقة خطية بين متغيرين .

(٩ – ٨) تحليل الانحدار حين يكون هناك أكثر من قيمة ص لكل قيمة س .

فى الأمثلة والتمارين السابقة كانت التجربة تحدد قيمة عشوائية واحدة ص لكل قيمة ثابتة س . إلا أن بعض الأبحاث تفضل تصميم التجربة بحيث نحصل منها على أكثر من قيمة عشوائية من المتغير ص لكل قيمة من قيم المتغير س ، خاصة وأن هذا التصميم يوفر لنا الفرصة للحكم على جودة العلاقة المفروضة بين المتغيرين كما سنرى فى البند (٩ – ٩) .

فإذا كان هناك ك من القيم السينية $س_{١}$ ، $س_{٢}$ ، ... ، $س_{ك}$ يكون لدينا ك من المجموعات الصادية المناظرة لها ، وتتخذ البيانات فى هذه الحالة الصورة المبينة بالجدول (٩ – ٥) والتى تشبه الصورة التى يتخذها تحليل التباين للتجارب ذوات العامل الواحد .

الجدول (٩ – ٥)

$س_{١}$	$س_{٢}$	...	$س_{و}$	...	$س_{ك}$
$ص_{١١}$	$ص_{٢١}$	...	$ص_{١و}$	...	$ص_{١ك}$
$ص_{١٢}$	$ص_{٢٢}$	...	$ص_{٢و}$	...	$ص_{٢ك}$
...	...	...	...	...	...
...	...	...	...	...	...
$ص_{ر١}$	$ص_{ر٢}$	...	$ص_{رو}$	...	$ص_{رك}$
...	...	...	...	...	...
...	...	...	...	...	...
$ص_{ن١}$	$ص_{ن٢}$	...	$ص_{ن_{و}و}$	...	$ص_{ن_{ك}ك}$

وإذا تبنينا نفس الافتراضات الثلاثة للانحدار الخطى فإن أسلوب التحليل يسير على نفس النمط السابق تقديمه مع بعض التعديلات التى يقتضيها الوضع الجديد للبيانات كما يتبين من المثال الآتى .

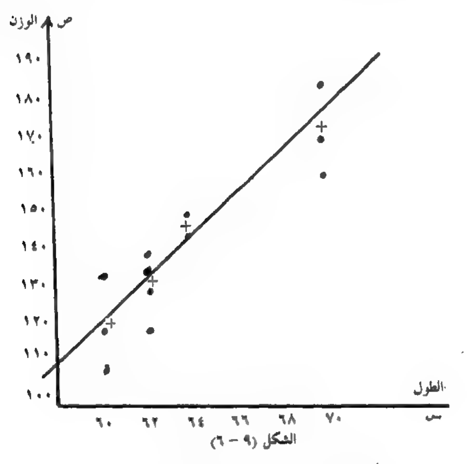
مثال (٩ – ٨) :

فى دراسة عن توزيع الأوزان لمجتمع من الرجال وعلاقة هذا التوزيع بالأطوال ، قسم مجتمع الرجال من حيث الطول إلى ٤ أقسام تتساوى فيها الأطوال بالتقريب وتمثلها الأطوال ٦٠ ، ٦٢ ، ٦٤ ، ٧٠ . وفى كل من هذه الأطوال اختير عدد من الرجال عشوائيا وقيست أوزانهم وسجلت المقاييس فى الجزء العلوى من الجدول (٩ – ٦) الآتى :

الجدول (٩ – ٦)

	الأطوال				
	$س_1$	$س_2$	$س_3$	$س_4$	
	٦٠	٦٢	٦٤	٧٠	
الأوزان	١١٠	١٢٠	١٥٠	١٧٠	
	١٣٥	١٤٠	١٤٥	١٨٥	
	١٢٠	١٣٠		١٦٠	
		١٣٥			
$ن_ى$	٣	٤	٢	٣	ن = ١٢
$مج_ر ص_{رى}$	٣٦٥	٥٢٥	٢٩٥	٥١٥	مجـ مجـ ص = ١٧٠٠
$مج_ر ص^2_{رى}$	٤٤٧٢٥	٦٩١٢٥	٤٣٥٢٥	٨٨٧٢٥	مجـ مجـ $ص^2$ = ٢٤٦١٠٠
$ن_ى س_ى$	١٨٠	٢٤٨	١٢٨	٢١٠	مجـ مجـ س = ٧٦٦
$ن_ى س^2_ى$	١٠٨٠٠	١٥٣٧٦	٨١٩٢	١٤٧٠٠	مجـ مجـ $س^2$ = ٤٩٠٦٨
$س_ى مج_ر ص_{رى}$	٢١٩٠٠	٣٢٥٥٠	١٨٨٨٠	٣٦٠٥٠	مجـ مجـ س ص = ١٠٩٣٨٠

نعبر عن هذه البيانات هندسيا كما فى الشكل (٩ – ٦) الذى يعرض شكل الانتشار ويشتمل على ١٢ نقطة تشترك بعضها فى الإحداثيات السينية وكل نقطة تعبر عن طول ووزن أحد الرجال . أما النقط المشار إليها بالعلامة + فتمثل متوسطات المجموعات الصادية عند القيم السينية المناظرة أى تمثل النقط ($س_و$ ، $\bar{ص}_و$) حيث و = ١ ، ٢ ، ٣ ، ٤ .



الشكل (٩ – ٦)

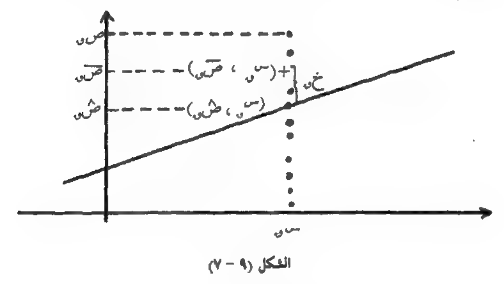
شكل الانتشار وخط الانحدار لبيانات المثال (٩ – ٨)

المطلوب فى هذا المثال إيجاد معادلة انحدار ص على س واختبار دلالة هذا الانحدار .

إيجاد معادلة الانحدار :

فى المثال (٩ – ٧) حيث كان لدينا قيمة واحدة ص لكل قيمة س كان بحثنا يهدف إلى معرفة ماإذا كانت القيم الصادية $ص_١$ ، $ص_٢$ ،... تقع على خط مستقيم هو خط الانحدار،

أما فى المثال (٩ — ٨) حيث لدينا مجموعة من القيم الصادية لكل قيمة س فان بحثنا يهدف فى هذه الحالة إلى معرفة ما إذا كانت المتوسطات ص̄١ ، ص̄٢ ، ... للمجموعات الصادية تقع على خط مستقيم . وعلى ذلك فإنه حسب مبدأ المربعات الصغرى — راجع البند (٩ — ٣) — يكون خط الانحدار هو ذلك الذى تختار ثوابته بحيث تجعل الدالة مج (ص̄ى — ص̂ى)٢ نهاية صغرى ، حيث ص̂ى هو الإحداثى الصادى للنقطة الواقعة على خط الانحدار والتى إحداثيها السينى سى — انظر الشكل (٩ — ٧) .



الشكل (٩ — ٧)

ويهمنا أن نشير إلى أننا إذا وضعنا بدلا من كل قيمة صى ى فى أى عمود ه الوسط الحسابى ص̄ى للصادات فى هذا العمود فإن معادلة الانحدار التى تنتج تكون هى بذاتها معادلة انحدار ص على س لأن هذا التغيير لا يؤثر فى المجاميع أو مجاميع المربعات أو مجاميع حواصل الضرب ، وهذا ما يمكن بيانه جبريا . وهذه الحقيقة تعنى أنه إذا وقعت متوسطات الأعمدة على خط مستقيم فإن هذا الخط ينطبق على خط انحدار ص على س .

لايجاد معامل الانحدار ب نستخدم الصيغة (٥) التى تأخذ فى هذه الحالة الشكل الآتى :

$$ب = \frac{\text{مج مج س ص} - \frac{(\text{مج مج س})(\text{مج مج ص})}{ن}}{\text{مج مج س}^{٢} - \frac{(\text{مج مج س})^{٢}}{ن}}$$

وفى هذا المثال نجد أن :

$$ب = \frac{١٠٩٣٨٠ - \frac{٧٦٦ \times ١٧٠٠}{١٢}}{٤٩٠٦٨ - \frac{٧٦٦^{٢}}{١٢}} = \frac{٨٦٣,٣٤}{١٧١,٦٦٧} = ٥,٠٢٩٢$$

$$\overline{س} = \frac{١}{ن} \text{ مج مج س} = \frac{٧٦٦}{١٢} = ٦٣,٨٣٣٣$$

$$\overline{ص} = \frac{١}{ن} \text{ مج مج ص} = \frac{١٧٠٠}{١٢} = ١٤١,٦٦٦٧$$

∴ معادلة انحدار ص على س مقربة إلى ٣ خانات عشرية هى :

$$\hat{ص} = ١٤١,٦٦٧ + ٥,٠٢٩ \text{ (س} - ٦٣,٨٣٣)$$

$$\text{أى } \hat{ص} = - ١٧٩,٣٤٩ + ٥,٠٢٩ \text{ س}$$

اختبار وجود علاقة خطية :

على فرض توفر شروط الانحدار يمكننا اختبار وجود علاقة خطية بين المتغيرين أى اختبار الفرض الصفرى $\beta = ٠$ ضد الفرض $\beta \neq ٠$ باستخدام اختبار ت بالصيغة (١١) أو اختبار ف بالصيغة العامة (٢١) أو بالصيغة المكافئة (٢٤) . وسنستخدم هنا الصيغة العامة وهى :

$$ف = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}} \qquad \text{بدرجتى حرية ١ ، ن} - ٢$$

من بيانات المثال نجد مايلي :

الاختلاف الكلي $= م\ م\ (ص) = مجـ\ مجـ\ ص^{٢} - \frac{(مجـ\ مجـ\ ص)^{٢}}{ن}$

$= ٢٤٦١٠٠ - \frac{١٧٠٠^{٢}}{١٢} = ٥٢٦٦,٦٧$

الاختلاف المفسر $= \frac{[م\ حـ\ (س\ ،\ ص)]^{٢}}{م\ م\ (س)}$

$= \frac{\left[١٠٩٣٨٠ - \frac{٧٦٦ \times ١٧٠٠}{١٢}\right]^{٢}}{٤٩٠٦٨ - \frac{٧٦٦^{٢}}{١٢}} = \frac{(٨٦٣,٣٤)^{٢}}{١٧١,٦٦٧} = ٤٣٤١,٨٧٠٨$

∴ الاختلاف غير المفسر = ٥٢٦٦,٦٧ − ٤٣٤١,٨٧٠٨ = ٩٢٤,٧٩٩٢

وبذلك نحصل على جدول التباين الآتي :

الجدول (٩ – ٧)

مصدر التباين	م م	د ح	ط م	ف ى
الاختلاف المفسر	٤٣٤١,٨٧٠٨	١	٤٣٤١,٨٧٠٨	٤٦,٩٥**
الاختلاف غير المفسر	٩٢٤,٧٩٩٢	١٠	٩٢,٤٧٩٩	
الاختلاف الكلي	٥٢٦٦,٦٧٠٠	١١		

بما أن ٤٦,٩٥ أكبر بكثير من القيمة الحرجة $ف_{٠,٠١[١\ ،\ ١٠]} = ١٠,٠٠$
نرفض الفرض الصفري $\beta = ٠$ عند مستوى عالي من الدلالة ونحكم بوجود علاقة خطية بين طول الرجل ووزنه .

(٩ – ٩) اختبار جودة العلاقة الخطية :

نعلم أننا إذا قبلنا الفرض الصفرى β = ٠ فإننا نحكم بعدم وجود علاقة خطية بين المتغيرين ولا نحتاج حينئذ إلى إجراء أى اختبارات أخرى تتعلق بخطية هذه العلاقة . أما إذا رفضنا هذا الفرض فإننا نحكم بوجود علاقة خطية بين المتغيرين لأن الانحدار الخطى يكون قد فسر جزءا ذا دلالة من الاختلاف الكلى . غير أن هذا لا يعنى أن العلاقة الخطية هى أحسن علاقة تصف العلاقة الحقيقية بين المتغيرين ، فقد تكون هناك علاقة أخرى مثل ص = α + β_{1} س + β_{2} س^{2} تفضل العلاقة الخطية فى ذلك . ويستلزم الأمر هنا الاستمرار فى تحليل البيانات لتقدير درجة جودة العلاقة الخطية ، وذلك بتقييم الاختلاف غير المفسر الذى يعبر عن الانحراف عن الخطية ، فإذا كان هذا الاختلاف غير ذى دلالة أى يرجع إلى العوامل العشوائية فإن العلاقة الخطية تكون هى أحسن العلاقات ، أما إذا كان هذا الاختلاف ذا دلالة فإن العلاقة الخطية لا تكون هى العلاقة المثلى بين المتغيرين .

ولكن كيف نقيّم الاختلاف المعبر عن الانحراف عن الخطية ؟ إننا نحتاج هنا إلى مقياس نقيس به دلالة هذا الانحراف . ولتحقيق هذا الغرض ينبغى تصميم تجربة نحصل منها على بيانات بحيث يناظر كل قيمة من القيم السينية مجموعة من القيم الصادية كما فى المثال (٩ – ٨) السابق ، لأن وجود هذه المجموعات الصادية يتيح لنا فرصة إيجاد الاختلاف داخل هذه المجموعات وهو الذى يعبر عن خطأ التجريب ، وبالتالى يمكن اتخاذه معيارا لمدى دلالة الانحراف عن الخطية .

وعلى ذلك ، وعلى فرض أن لدينا بيانات من مثل هذه التجربة نبدأ بخطوة هامة هى تحليل التباين للمجموعات الصادية ، أى فصل الاختلاف الكلى فى القيم الصادية (كالمعتاد) إلى مركبتين مستقلتين تعبر الأولى عن الاختلاف بين المجموعات وتعبر الثانية عن الاختلاف داخل المجموعات (خطأ التجريب) ، وهذه الخطوة تتخذ الشكل المعتاد الآتى :

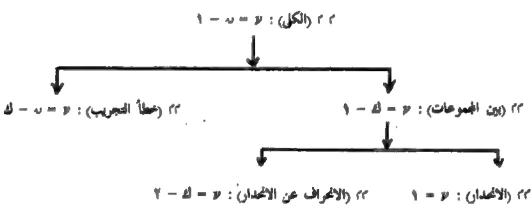
م م (ص) = م م (بين المجموعات) + م م (داخل المجموعات)
بدرجات حرية ن − ١ ، ك − ١ ، ن − ك على الترتيب .

وجدير بالإشارة هنا إلى أن المتغير المستقل س في الانحدار هو متغير كمى بينما المتغير المستقل (عامل التجريب) فى تحليل التباين هو متغير نوعى . وحين نجرى تحليل التباين للانحدار نعتبر أن القيم العددية للمتغير س هى مجرد إشارات تعبر عن مستويات عامل التجريب .

ولما كان الهدف من تحليل الانحدار اختبار ما إذا كانت متوسطات المجموعات الصادية تقع على خط مستقيم فإن الاختلاف الذى ينبغى تحليله هو الاختلاف **بين المجموعات** . وعلى ذلك فإن الخطوة الثانية هى تحليل الاختلاف بين المجموعات إلى مركبتين مستقلتين تعبر الأولى عن الاختلاف المفسر : م م (الانحدار الخطى) وتعبر الثانية عن الاختلاف غير المفسر : م م (الانحراف عن الانحدار الخطى) وهذه الخطوة تتخذ الشكل الآتى :

م م (بين المجموعات) = م م (الانحدار الخطى) + م م (الانحراف عن خط الانحدار)
بدرجات حرية ك − ١ ، ١ ، ك − ٢ على الترتيب .

الشكل (٩ − ٨) الآتى يلخص خطوتى التحليل السابق ذكرهما .



الشكل (٩ − ٨) : خطوتا اختبار جودة الانحدار الخطى

كما أن الجدول (٩ − ٨) الآتى هو جدول التباين لعمليتى التحليل .

الجدول (٩ − ٨)

مصدر التباين	مجموع المربعات	درجات الحرية
بين المجموعات	مج نى (ص̄ى − ص̄)٢	ك − ١
{ الانحراف الخطى	{ مج نى (ص̂ى − ص̄)٢	١
{ الانحراف عن خط الانحدار	{ مج نى (ص̄ى − ص̂ى)٢	ك − ٢
خطأ التجريب (داخل المجموعات)	مج مج (صرى − ص̄ى)٢	ن − ك .
الكلى	مج مج (صرى − ص̄)٢	ن − ١

من هذا التحليل نستطيع أن نختبر ثلاثة أمور هى :

(أولا) اختبار دلالة الاختلاف المشاهد بين المجموعات :

أى اختبار ما إذا كانت متوسطات المجموعات الصادية تختلف باختلاف القيم السينية . والاختبار الذى يصلح لذلك هو اختبار ف المعتاد حيث

ف = (التباين بين المجموعات) / (التباين داخل المجموعات) بدرجتى حرية ك − ١ ، ن − ك (٢٥)

والفرض الصفرى هنا هو أن المتوسطات متساوية جميعها . فإذا قبلنا هذا الفرض نحكم بعدم وجود أى علاقة بين التغير فى س والتغير فى ص ويتوقف البحث عند هذا الحد . أما إذا رفضنا الفرض الصفرى فنحكم بأن التغير فى س يؤثر فى التغير فى ص وينبغى حينئذ أن نستمر فى البحث بحسب الخطوة الثانية .

(ثانيا) بحث العلاقة الخطية بين المتغيرين :

(ا) اختبار وجود علاقة خطية :

أى اختبار الفرض الصفرى $\beta = 0$ ضد الفرض $\beta \neq 0$ وكما فعلنا فى المثال (٩ — ٨) نستخدم الصيغة (٢١) وهى :

ف = التباين المفسر / التباين غير المفسر

مع ملاحظة أن الاختلاف غير المفسر هو ذلك الاختلاف الذى لا يعتمد على المتغير س وهو يشمل الاختلاف الناشىء عن الانحراف عن خط الانحدار كما يشمل الاختلاف الناشىء عن خطأ التجريب ، وعلى ذلك فإن التعويض فى هذه الصيغة من بيانات الجدول (٩ — ٨) يكون كالآتى :

ف = [م ٢ (الانحدار الخطى) ÷ ١] / [[م ٢ (الانحراف عن الخطية) + م ٢ (خطأ التجريب)] ÷ (ن—٢)] (٢٦)

بدرجتى حرية ١ ، ن — ٢ .

وكما سبق القول فى (أولا) ليس هناك ما يدعو للقيام بهذا الاختبار أو بالاختبار الذى سيرد فى (ب) إلا إذا ثبت من الخطوة السابقة دلالة الاختلاف بين المجموعات الصادية .

(ب) اختبار جودة العلاقة الخطية :

أى اختبار دلالة الانحراف عن خط الانحدار مقدرا بمجموع مربعات الانحرافات $\bar{ص}_س - \hat{ص}_س$ ، أو بمعنى آخر اختبار ما إذا كانت متوسطات المجموعات الصادية تقع على خط الانحدار أو قريبة منه أم تنتشر بعيدة عنه أبعادا جوهرية . والاختبار الذى يصلح لذلك هو اختبار ف بالصورة

$$\text{ف} = \frac{\text{تباين الانحراف عن الخطية}}{\text{تباين خطأ التجريب}} \qquad (٢٧)$$

بدرجتى حرية ك − ٢ ، ن − ك .

فى هذا الاختبار يلعب التباين المقدر لخطأ التجريب دورا رئيسيا كمعيار يقاس بالنسبة إليه الانحراف عن الخطية ، وهذا هو السبب فى ضرورة أن تصمم التجارب التى تهدف إلى قياس جودة العلاقات الخطية بحيث يكون لكل قيمة س قيمتان أو أكثر من قيم ص وإلا ما استطعنا الحصول على هذا المعيار .

ملاحظة :

إذا وجدنا من الخطوة (أ) أن الانحدار الخطى ذو دلالة ووجدنا من الخطوة (ب) أن الانحراف عن الخطية غير ذى دلالة ، نحكم بأن العلاقة الخطية هى علاقة جيدة وتصف بجدارة العلاقة الحقيقية بين المتغيرين . أما إذا وجدنا أن كلا من الانحدار الخطى والانحراف عن الخطية ذو دلالة فنحكم بأنه بالرغم من وجود علاقة خطية بين المتغيرين إلا أنها ليست أحسن العلاقات التى تعبر عن حقيقة العلاقة بينهما وعلينا إذا أردنا أن نبحث عن علاقة أفضل .

مثال (٩ − ٩) :

ابحث خطية العلاقة بين الطول والوزن مستخدما بيانات المثال (٩ − ٨) .

الحل :

(أولا) نبدأ بتحليل التباين للمتغير ص للحكم على دلالة الاختلاف بين متوسطات المجموعات الصادية .

م م (الكلى) = ٥٢٦٦,٦٧ سبق إيجاده بالمثال (٩−٨)

$$\text{م م (بين المجموعات)} = \frac{{٣٦٥}^{٢}}{٣} + \frac{{٥٢٥}^{٢}}{٤} + \frac{{٢٩٥}^{٢}}{٢} + \frac{{٥١٥}^{٢}}{٣} - \frac{{١٧٠٠}^{٢}}{١٢}$$

= ٤٤٠٢,٠٨ $\nu = ٣$

∴ م م (خطأ التجريب) = ٥٢٦٦,٦٧ − ٤٤٠٢,٠٨ = ٨٦٤,٥٩ $\nu = ٨$

من (٢٥) : ف $= \frac{٤٤٠٢,٠٨ \div ٣}{٨٦٤,٥٩ \div ٨} = ١٣,٥٧٧^{**}$

وبما أن ف $_{٠,٠١ [٣ ، ٨]}$ = ٧,٥٩ نرفض الفرض الصفرى عن تساوى المتوسطات عند مستوى الدلالة ٠,٠١ ، ونحكم بأن أوزان الرجال ليست مستقلة عن أطوالهم . ومادام الأمر كذلك نستمر فى التحليل .

(ثانيا) نقوم بتحليل الانحدار للاختلاف بين المجموعات كما يلى :

م م (الانحدار الخطى) = ٤٣٤١,٨٧٠٨ سبق إيجاده بالمثال (٩−٨) $\nu = ١$

∴ م م (الانحراف عن الخطية) = م م (بين المجموعات) − م م (الانحدار الخطى)

= ٤٤٠٢,٠٨ − ٤٣٤١,٨٧٠٨

= ٦٠,٢٠٩٢ $\nu = ٢$

بضم هذه النتائج إلى نتائج الخطوة السابقة ينتج جدول التباين الآتى :

الجدول (٩ - ٩)

مصدر التباين	م م	د ح	ط م
بين المجموعات	٤٤٠٢,٠٨	٣	١٤٦٧,٣٦
الانحدار الخطى	٤٣٤١,٨٧٠٨	١	٤٣٤١,٨٧٠٨
الانحراف عن الخطية	٦٠,٢٠٩٢	٢	٣٠,١٠٤٦
خطأ التجريب	٨٦٤,٥٩	٨	١٠٨,٠٧٣٨
الكلى	٥٢٦٦,٦٧	١١	

(أ) اختبار وجود علاقة خطية

من (٢٦) : $$\text{ف}_{\text{ى}} = \frac{١ \div ٤٣٤١,٨٧٠٨}{١٠ \div (٨٦٤,٥٩ + ٦٠,٢٠٩٢)} = ٤٦,٩٥^{**}$$

بما أن $\text{ف}_{٠,٠١\,[١ ، ١٠]}$ = ١٠,٠٠ انرفض الفرض الصفرى أن β = ٠ ونحكم بوجود علاقة خطية بين أطوال الرجال وأوزانهم .

(ب) اختبار دلالة الانحراف عن خط الانحدار

من (٢٧) : $$\text{ف}_{\text{ى}} = \frac{٣٠,١٠٤٦}{١٠٨,٠٧٣٨} < ١$$

وإذن نقبل الفرض الصفرى بعدم وجود انحراف ذى دلالة عن خط الانحدار .

الخلاصة :

من هذه التجربة نستخلص أن العلاقة الخطية التى تمثلها معادلة الانحدار ص̂ = − ١٧٩,٣٥ + ٥,٠٢٩ س تعبر بجدارة عن العلاقة الحقيقية بين أطوال وأوزان الرجال . ومع ملاحظة أن معامل التحديد

$$ر^٢ = \frac{[م ح (س ، ص)]^٢}{م م (س) . م م (ص)}$$

$$= \frac{(٨٦٣,٣٤)^٢}{١٧١,٦٦٧ \times ٥٢٦٦,٦٧} = ٠,٨٢٤٤$$

نرى أن العلاقة الخطية تفسر حوالى ٨٢,٤٪ من الاختلاف الكلى فى قيم المتغير ص ، وهذا يدعم القول بخطية العلاقة بين المتغيرين .

(٩ − ١٠) ملاحظات عن افتراضات الانحدار :

إن صحة الاستنتاجات الإحصائية تتوقف على مدى انطباق الافتراضات التى وضعت فى تعريف النموذج الذى بنيت عليه الدراسة على الموقف التجريبى الذى تتناوله هذه الدراسة . ولذلك ينبغى أن نتفهم جيدا ما تعنيه هذه الافتراضات وما تتطلبه من شروط إجرائية تحسبا من العواقب التى قد تنجم عن وجود تناقضات ذات بال بين الافتراضات الموضوعة والظروف الفعلية لعملية التجريب . كما ينبغى أن نكون على استعداد لتغيير افتراض أو تعديله إذا اتضح لنا عدم إمكانية تحققه عمليا ولو بشىء من التقريب ، على أن نراعى ما يستلزمه هذا التغيير أو التعديل بالنسبة لما نقدمه من استنتاجات .

لنعتبر الافتراضات الثلاثة التى استخدمناها فى الانحدار الخطى البسيط ولنبدأ بالافتراض الأول . إن هذا الافتراض يتضمن أن يكون المتغير س متغيرا غير عشوائى

وهذا يعنى أن الباحث يتحكم تجريبيا فى هذا المتغير ويستطيع تسجيل القيم التى يدخلها فى بياناته بدقة تامة . غير أنه فى كثير من المواقف التجريبية لا يتحقق هذا الافتراض فقد يكون الباحث مهتما بالحصول على بيانات عن المتغير س دون أن يكون متحكما فيه أى بصرف النظر عن كون هذا المتغير عشوائيا أو غير عشوائى مادامت هذه البيانات تعبر فى نظر الباحث عن قيم نموذجية لهذا المتغير . وفى بعض الدراسات يضطر الباحث إلى اختيار مشاهدات تقع فى مدى معين محدد من قبل أو يأخذ منها قيما معينة محددة مسبقا . فى مثل هذه الحال ينبغى للباحث أن يستبدل بهذا الافتراض افتراضا آخر يتلاءم مع الموقف الذى يتناوله ، كأن يضع الافتراض «المتغير س قيمه مشروطة » وهذا الافتراض الجديد لا يؤثر فى سلامة استخدام طريقة المربعات الصغرى إلا أن الاستنتاجات الإحصائية التى نخرج بها عن المجتمع الذى ندرسه ينبغى أن تكون مشروطة بالنسبة للمتغير س أى لا يجوز تعميمها لقيم غير ممثلة فى البيانات التى استخدمت فى الدراسة أو ليس لها نفس خصائصها .

أما الافتراض الثانى فيعنى أن الصيغة الحقيقية للعلاقة بين المتغيرين هى الصيغة الخطية ، وعلى هذا لا تكون استنتاجاتنا صحيحة إلا إذا كان لدينا ما يضمن سلامة هذه الصيغة ، ويساعدنا فى ذلك الأسلوب المذكور فى نتيجة البند (٩ — ٥) لاختبار خطية هذه العلاقة كما يساعدنا الأسلوب المبين بالبند (٩ — ٩) لاختبار دلالة الانحراف عن خط الانحدار . فإذا تبين لنا عدم انطباق هذه الصيغة فى موقف ما ينبغى أن نبحث عن صيغة أخرى تناسبه .

وبالنسبة للافتراض الثالث فهو يعنى أنه عند أى قيمة ثابتة من المتغير س تتصف القيم التى يأخذها المتغير ص بما يلى :

(١) قيم ص مستقلة ، أى أن صغر أو كبر الخطأ العشوائى فى أحدها لا يؤثر فى مقدار الخطأ فى القيم الأخرى . وهذه الصفة يمكن تحقيقها عمليا بإحكام عملية التجريب وخاصة فيما يتعلق بعشوائية العينة .

(٢) قيم ص لها تباين ثابت σ^2 ، وهذه الصفة تتحقق تلقائيا إذا كانت هذه القيم هى مشاهدات مستقلة من نفس المجتمع .

(٣) قيم ص تتوزع توزيعا معتدلا .

وجدير بالإشارة هنا إلى أنه بصفة عامة تكون النماذج المستخدمة بما يصاحبها من فروض هى نماذج نظرية قلما تتحقق فى الواقع العملى إلا على وجه التقريب . ونحن إذ نستخدم هذه النماذج نتعلق بالأمل فى ألا تؤدى بنا هذه الحقيقة إلى وجود فروق كبيرة بين التقديرات التى نحصل عليها منها وبين القيم الحقيقية للبارامترات التى نبحث عنها ، كما نتعلق بالأمل بأن تكون الإجراءات التى اتخذناها فى عملية التقدير هى إجراءات مناسبة حتى إذا كان الموقف الذى نتناوله لا يحقق الفروض تحقيقا تاما .

(٩ - ١١) استخدامات الانحدار :

لعله من المناسب الآن أن نبرز الأغراض التى يستخدم الانحدار الخطى البسيط من أجلها . ومن الدراسة التى مرت بنا فى هذا الفصل يمكننا تلخيص هذه الأغراض فيما يلى :

(١) التنبؤ بقيم متغير عشوائى ص بمعلومية متغير رياضى س مع تقدير درجة دقة هذا التنبؤ . هذا مع ملاحظة أن خط الانحدار هو خط مستقيم يمتد بغير نهاية من الطرفين . ولكننا فى عملية التنبؤ لا يجوز التنبؤ بقيم صادية مناظرة لقيم سينية تخرج كثيرا عن مدى القيم التى استخدمت فى إنشاء هذا الخط إلا إذا كان لدينا ما يبرر ذلك . فمثلا إذا أوجدنا خط انحدار لأطوال الذكور بين العمر ١٠ والعمر ١٥ فمن الخطأ استخدام هذا الخط للتنبؤ بأطوال الرجال فى أعمار فوق العشرين لأن الطول يتوقف عند بلوغ سن النضج .

(٢) دراسة طبيعة العلاقة بين متغيرين س ، ص وشكل المنحنى أو الدالة التى تعبر عن هذه العلاقة .

(٣) تفسير بعض الاختلاف فى قيم المتغير ص بدلالة الاختلاف فى قيم س على أساس اتخاذ المتغير س كضابط إحصائى ststistical control .

(٤) دراسة ما إذا كان التغير فى س **يسبب** ولو جزئيا التغير فى ص ، مع ملاحظة أن دراسة السببية تحتاج إلى أكثر من إثبات وجود علاقة ذات دلالة بين المتغيرين ، إذ أن هذه العلاقة قد تكون ناشئة عن وجود متغيرات أو عوامل أخرى تؤثر فى المتغيرين س ، ص معا . فمثلا فى الأحياء المزدحمة من بعض المدن الكبيرة نجد علاقة ذات دلالة بين ازدحام السكان والتدرن الرئوى (السل) فهل هذا يعنى أن الازدحام سبب فى الإصابة بالتدرن الرئوى ؟ لا نستطيع الإجابة بالإيجاب أو النفى عن هذا السؤال إذا لاحظنا أنه فى الأحياء المزدحمة بصفة عامة يكون مستوى الحياة منخفضا وبالتالى يكون هناك سوء تغذية للسكان ، وقد يكون سوء التغذية هو السبب الحقيقى لهذا المرض . إن تقرير السببية أمر متروك للباحث يفتى فيه بما لديه من خبرة ومعلومات عن المتغيرات التى يتناولها مستعينا بما يستخلصه من التحليل الاحصائى .

(٥) يستخدم الانحدار أيضا فى أنواع أخرى من التحليل الإحصائى منها تحليل التغاير حيث يكون الهدف معرفة مدى تأثير عامل نوعى على متغير عددى ص بعد استبعاد أثر متغير عددى س مرتبط بالمتغير ص ، وهذا ما سنتناوله فى فصل لاحق .

تمارين (٩ – ٢)

(اعتبر أن افتراضات الانحدار متوفرة) .

(١) فى تجربة عن تأثر طاقة التمثيل الغذائى بدرجة الحرارة على نوع من الطيور فى فترة ضوئية ثابتة مدتها ١٠ ساعات ، اختيرت أربع درجات حرارة هى ٥ ، ١٠ ، ١٥ ، ٢٠ وعرضت خمسة طيور لكل من هذه الدرجات ثم حسبت مقادير طاقة التمثيل الغذائى لكل من هذه الطيور وسجلت بالجدول الآتى . أوجد معادلة

الانحدار الخطى لطاقة التمثيل الغذائى على درجة الحرارة واختبر جودة العلاقة الخطية بين هذين المتغيرين .

	درجات الحرارة (س)			
	٥	١٠	١٥	٢٠
طاقة التمثيل الغذائى (ص)	٢٣,٢	٢٤,٢	١٨,٣	١٥,٨
	٢٣,٥	٢٤,٢	١٨,٨	١٥,٢
	٢٣,٤	٢٤,١	١٨,٦	١٥,٦
	٢٣,١	٢٤,٣	١٨,٧	١٥,٣
	٢٣,٦	٢٤,٢	١٨,١	١٥,١

لكل من العينات الثلاث الآتية أوجد معادلة الانحدار الخطى واختبر جودة العلاقة الخطية :

(٢)

	س				
	٠	١	٢	٤	٦
ص	١-	٠	١	٨	١١
	١	٢	٥	١٢	١٣

(٣)

	س			
	١	٢	٣	٤
ص	٠	٥	٩	١٠
	٢	٧	١٣	١٤

(٤)

	س					
	٢	٥	١٠	١٥	٢٠	٢٥
ص	٦٠	١٧١	٢٤٢	٢٩٨	٣٤٥	٣٨٣
	٦٤	١٧٦	٢٤٥	٣٠٤	٣٥١	٣٨٨

الفصل العاشر

الارتباط الخطى البسيط

SIMPLE LINEAR CORRELATION

(١٠ - ١) الانحدار الخطى والارتباط الخطى :

في دراسة الانحدار الخطى لمتغير حقيقي ص على متغير آخر س فرضنا أن العلاقة بينهما على الصورة $\mu_{ص \cdot س} = \alpha + \beta$ س واستخرجنا من العينة أحسن تقديرين ا ، ب للبارامترين المجهولين α ، β في ضوء مبدأ المربعات الصغرى ، ومن ثم أوجدنا معادلة $\hat{ص}$ = ا + ب س تمكننا من التنبؤ بأحسن قيمة للمتغير ص عند قيمة معطاة للمتغير س ، واستخدمنا تحليل التباين لتفسير الانحدار ، كما خرجنا ببعض الاستنتاجات الإحصائية في صورة اختبارات دلالة وفترات ثقة . وهذا كله يدخل تحت الموضوع المسمى **بتحليل الانحدار** . ويقترن بهذا الموضوع موضوع آخر يسمى **تحليل الارتباط** وهو يهتم بالبحث عن عدد نقيس به درجة الاعتماد المتبادل بين المتغيرين ودقة العلاقة المفروضة بينهما ، فإذا فرضنا أن العلاقة بين المتغيرين خطية يكون اهتمامنا منصباً على إيجاد عدد أو مقياس يعبر عن درجة جودة العلاقة الخطية في وصف العلاقة الحقيقية بين المتغيرين واختبار دلالة هذا المقياس .

(١٠ - ٢) افتراضات الارتباط الخطى البسيط :

تقتضي دراسة الارتباط بين متغيرين حقيقيين س ، ص وضع افتراضات يختلف بعضها عن تلك التي وضعت لدراسة الانحدار . وسنضع هنا الافتراضات الآتية :

الافتراض الأول :

« كل من المتغيرين س ، ص هو متغير عشوائي » .

فمثلا قد يعبر المتغيران عن طول ذراع الإنسان وطول رجله ، أو عن عمر الزوج وعمر الزوجة عند الزواج ، أو عن وزن الدجاجة وعدد البيض الذى تنتجه أو عن درجة الرياضيات ودرجة الفيزياء لمجموعة من الطلاب ..

الافتراض الثانى :

« **إن العلاقة بين المتغيرين س ، ص هى علاقة خطية** » بمعنى أن متوسط قيم ص المناظرة لقيمة معينة س يأخذ الصورة

$$\mu_{ص|س} = \alpha + \beta س \qquad (1)$$

حيث α ، β بارامتران مجهولان .

الافتراض الثالث :

(أ) للمتغير س توزيع معتدل

(ب) عند أى قيمة ثابتة س يكون للمتغير ص توزيع معتدل متوسطه $\alpha + \beta$س وتباينه عدد ثابت مجهول σ^2 لا يتوقف على س .

وقد يكون من المفيد أن نشير إلى أن هذه الافتراضات تكافىء رياضيا القول بأن التوزيع المشترك للمتغيرين س ، ص هو ذلك المسمى بالتوزيع المعتدل ذى المتغيرين bivariate normal distribution

(١٠ - ٣) معامل الارتباط العزمى (بيرسون) :

PRODUCT MOMENT CORRELATION COEFFICENT (PEARSON 1900)

تعريف :

إذا كان س ، ص متغيرين عشوائيين ورمزنا بالرمز $\sigma_{س}$ للانحراف المعيارى

للمتغير س وبالرمز $\sigma_{\text{ص}}$ للانحراف المعيارى للمتغير ص وبالرمز $\sigma_{\text{سص}}$ لتغاير (س ، ص) فإن العدد ر المعرف بالصيغة :

$$ر = \frac{\sigma_{\text{سص}}}{\sigma_{\text{س}} \cdot \sigma_{\text{ص}}}$$

يسمى بمعامل الارتباط العزمى بين س ، ص .

يمكن رياضيا إثبات ما يلى :

(١) لا تزيد القيمة المطلقة للعدد ر عن الواحد الصحيح ، أى أن

$$١ \geqslant ر \geqslant -١$$

(٢) إذا كان المتغيران س ، ص مستقلين فإن ر = صفر غير أن العكس ليس من الضرورى أن يكون صحيحا ، أى أنه إذا كان ر = ٠ فليس من الضرورى أن يكون المتغيران مستقلين بل قد يكون بينهما علاقة غير خطية . أما إذا كان التوزيع المشترك للمتغيرين س ، ص هو التوزيع المعتدل ذو المتغيرين فإن انعدام معامل الارتباط يستلزم استقلال المتغيرين .

(٣) تكون هناك علاقة دالية خطية بين المتغرين س ، ص :

ص = ا + ب س و س = اَ + بَ ص

إذا وإذا فقط كان معامل الارتباط ر يساوى ١ أو – ١ .

وهذه الخاصة تعنى أنه إذا كان هناك علاقة خطية بين س ، ص فإن ذلك ينعكس على قيمة ر فيجعلها مساوية للعدد ١ (ونقول حينئذ أن هناك ارتباط تام موجب بين المتغيرين) أو مساوية للعدد – ١ (ونقول حينئذ أن هناك ارتباط تام سالب) ، والعكس صحيح .

من هذا نرى أن معامل الارتباط العزمى ليس مقياسا للاعتماد المتبادل بين

المتغيرين بشكل عام وإنما هو مقياس لدرجة **الاعتماد الخطى** بينهما ، وينبغى أن نتذكر ذلك دائما .

ولتقدير قيمة ρ من عينة عشوائية :

$\{(س_1 ، ص_1) ، (س_2 ، ص_2) ، \dots ، (س_ن ، ص_ن)\}$ نستخدم المقياس الآتى الذى يتركب بكل بساطة من التقديرات غير المتحيزة للقيم التى يتركب منها ρ :

$$ر = \frac{ع_{س ص}}{ع_س \,.\, ع_ص} \qquad (٢)$$

حيث $ع_{س ص} = \frac{1}{ن - 1} \text{ مج } (س - \bar{س})(ص - \bar{ص})$ تغاير (س ، ص)

$$= \frac{1}{ن - 1} [\text{مج } س\, ص - (\text{مج } س)(\text{مج } ص) / ن] \qquad (٣)$$

، $ع_س^2 = \frac{1}{ن - 1} \text{ مج } (س - \bar{س})^2$ تباين (س)

$$= \frac{1}{ن - 1} [\text{مج } س^2 - (\text{مج } س)^2 / ن] \qquad (٤)$$

، $ع_ص^2 =$ تباين (ص)

ملاحظة (١)

من السهل إثبات أن (٢) يمكن أن تكتب على الصورة

$$ر = \frac{ن \text{ مج } س\, ص - (\text{مج } س)(\text{مج } ص)}{\sqrt{[ن \text{ مج } س^2 - (\text{مج } س)^2]\,[ن \text{ مج } ص^2 - (\text{مج } ص)^2]}} \qquad (٥)$$

وهذه الصيغة أفضل من الصيغة (٢) من الناحية الحسابية .

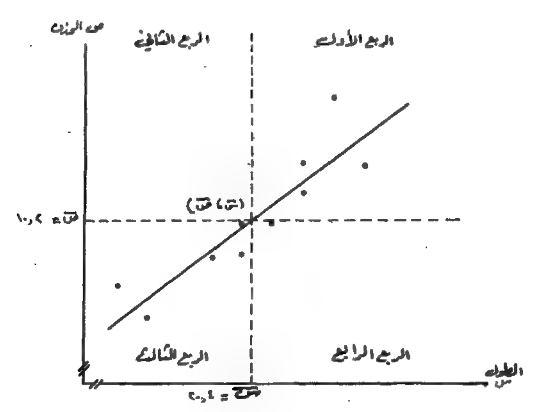
وقبل أن نوضح خصائص معامل الارتباط ر والحكمة في أخذه كمقياس للارتباط الخطى نتناول المثال الآتي :

مثال (١٠ – ١) :

أخذت عينة عشوائية من نوع معين من نبات البسلة وقيست أطوالها بالمليمتر (س) وأوزانها بالمليجرام (ص) فوجد ما يلى :

س :	١٧	٢٢	٢٠	٢٣	٢٤	٢٢	٢٠	١٩	٢١	١٦
ص :	٧	١٢	٩	١٤	١٢	١١	١٠	٩	١٠	٨

ارسم شكل الانتشار لاستيضاح خطية العلاقة بين المتغيرين ثم أوجد معامل الارتباط العزمى .



الشكل (١٠ – ١) شكل الانتشار لأطوال وأوزان عينة من ١٠ نباتات بسلة

يوحى شكل الانتشار بأن النقط تميل إلى أن تقع على خط مستقيم موجب الميل ، وهذا يشير مبدئيا إلى خطية العلاقة بين المتغيرين .

الجدول (١٠ – ١)

إيجاد معامل الارتباط العزمى من بيانات المثال (١٠ – ١)

س	ص	س ص	$س^٢$	$ص^٢$
١٧	٧	١١٩	٢٨٩	٤٩
٢٢	١٢	٢٦٤	٤٨٤	١٤٤
٢٠	٩	١٨٠	٤٠٠	٨١
٢٣	١٤	٣٢٢	٥٢٩	١٩٦
٢٤	١٢	٢٨٨	٥٧٦	١٤٤
٢٢	١١	٢٤٢	٤٨٤	١٢١
٢٠	١٠	٢٠٠	٤٠٠	١٠٠
١٩	٩	١٧١	٣٦١	٨١
٢١	١٠	٢١٠	٤٤١	١٠٠
١٦	٨	١٢٨	٢٥٦	٦٤
٢٠٤	١٠٢	٢١٢٤	٤٢٢٠	١٠٨٠

معامل الارتباط العزمى هو (بالتعويض في ٥)

$$ر = \frac{(١٠ \times ٢١٢٤) - (٢٠٤ \times ١٠٢)}{\sqrt{[١٠ \times ٤٢٢٠ - (٢٠٤)^{٢}][١٠ \times ١٠٨٠ - (١٠٢)^{٢}]}}$$

$$= \frac{٤٣٢}{\sqrt{٥٨٤ \times ٣٩٦}} = \frac{٤٣٢}{٤٨٠,٨٩٩}$$

= ٠,٨٩٨ تقريبا .

(١٠ – ٤) مميزات معامل الارتباط العزمي .

(أ) تركيب معامل الارتباط :

إن قدرة معامل الارتباط المعرف في (٢) على تقدير درجة العلاقة الخطية بين المتغيرين س ، ص تتبين من الدراسة الرياضية للنموذج الذى وضعت له الافتراضات المذكورة آنفاً . غير أننا نستطيع أن نرى ذلك بطريقة بسيطة كالآتي .

إذا تأملنا نقط مستقيم موجب الميل ، لاحظنا أنه كلما كان الإحداثي السيني لنقطة كبيراً كلما كان إحداثيها الصادى كبيراً أيضاً . أما إذا كان المستقيم سالب الميل فإنه كلما زاد الإحداثي السيني كلما نقص الإحداثي الصادى . وعلى ذلك فإن قياس خطية العلاقة بين المتغيرين تتطلب مقياساً حساساً لدرجة اقتران القيم الكبيرة للمتغير س بالقيم الكبيرة للمتغير ص إذا كانت العلاقة موجبة ، ولدرجة اقتران القيم الكبيرة للمتغير س بالقيم الصغيرة للمتغير ص إذا كانت العلاقة سالبة وهذه الحساسية نجدها في التغاير ع$_{س ص}$ ففي المثال (١٠ – ١) السابق ، إذا رسمنا الخط الرأسي س = $\bar{س}$ = ٢٠,٤ والخط الأفقي ص = $\bar{ص}$ = ١٠,٢ في شكل الانتشار لاحظنا أنه في أغلب الحالات بل في جميع الحالات ما عدا حالة واحدة هى حالة النقطة (٢١ ، ١٠) الواقعة في الربع الرابع ما يلي :

حين تكون س كبيرة (أكبر من الوسط الحسابى $\bar{س}$) فإن ص المناظرة تكون كبيرة أيضاً (أكبر من الوسط الحسابى $\bar{ص}$) وحين تكون س صغيرة (أصغر من $\bar{س}$) فإن ص المناظرة تكون صغيرة أيضاً (أصغر من $\bar{ص}$) . أى أن :

حين (س – $\bar{س}$) موجبة تكون (ص – $\bar{ص}$) موجبة وحين (س – $\bar{س}$) سالبة تكون (ص – $\bar{ص}$) سالبة وفي كلتا الحالتين تكون (س – $\bar{س}$) (ص – $\bar{ص}$) موجبة ويكون المجموع مج (س – $\bar{س}$) (ص – $\bar{ص}$) موجبا .

وبالتالى يكون التغاير $ع_{س ص}$ موجبا (إلا إذا كانت قيمة $(س - \bar{س})$ $(ص - \bar{ص})$ للنقطة (٢١ ، ١٠) كبيرة جداً وهذا لم يحدث) .

ويلاحظ أن قيمة $ع_{س ص}$ في هذا المثال كبيرة لأنها متوسط مجموع ٩ أعداد موجبة وعدد واحد سالب . ولو كانت النقطة التي في الربع الرابع قد وقعت في الربع الأول أو الثالث لزادت قيمة $ع_{س ص}$ وبالعكس لو كانت إحدى النقط الواقعة في الربعين الأول أو الثالث قد وقعت في الربع الثاني أو الرابع لنقصت قيمة $ع_{س ص}$

وعلى ذلك فإن التغاير يعكس أمرين هما : درجة جودة العلاقة الخطية واتجاه هذه العلاقة . ولهذا يرتكز معامل الارتباط ر المعرف في (٢) على التغاير $ع_{س ص}$. هذا مع ملاحظة أن هذا المعامل يكون موجباً أو سالباً بحسب كون العلاقة الخطية موجبة أو سالبة .

غير أن قيمة التغاير تعتمد على حجم العينة ن كما هو واضح من التعريف (٣) ، كما أنها تعتمد على وحدات القياس . ولكى يكون مقياس الارتباط عاماً ينبغى أن يكون مستقلا عن حجم العينة وعن وحدات القياس . ولذلك ينبغى تعديل التغاير ليحقق هذين الشرطين قبل أخذه كمقياس للارتباط ، وهذا ما يحققه قسمة التغاير $ع_{س ص}$ على حاصل ضرب الانحراف المعيارى $ع_{س}$ للمتغير س والانحراف المعيارى $ع_{ص}$ للمتغير ص . وهذا الإجراء يكافىء إيجاد تغاير س ، ص بعد وضع القيم في الصورة المعيارية $\frac{س - \bar{س}}{ع_{س}}$ ، $\frac{ص - \bar{ص}}{ع_{ص}}$

بهذا الإجراء يكون المقياس .

$ر = \frac{ع_{س ص}}{ع_{س} . ع_{ص}}$ هو مقياس مطلق (لا يعتمد على وحدات القياس) ولا يعتمد على حجم العينة .

(ب) من مزايا معامل الارتباط أنه متماثل في س ، ص بمعنى أن قيمته لا تتغير بوضع س ، ص كل مكان الآخر .

(حـ) قيمة معامل الارتباط ر مستقلة عن اختيار نقطة الأصل لأن كلا من التغاير والانحراف المعيارى يتمتع بهذه الصفة ، وهذا يعني أن قيمة ر لا تتغير بإضافة أو طرح عدد ثابت من جميع قيم س أو بإضافة أو طرح عدد ثابت من جميع قيم ص . ونستفيد من هذه الخاصية أحياناً في تبسيط حساب قيمة ر .

(ى) يمكن إثبات أن القيمة المطلقة للمقياس ر لا تزيد عن الواحد الصحيح . أى أن .

$$١ \geqslant ر \geqslant -١ \qquad (٦)$$

ولما كانت ر مقياساً لدرجة الارتباط الخطى بين متغيرين فإننا نقول إن المتغيرين غير مرتبطين خطياً إذا كانت ر = ٠ (وهذا لا يمنع من وجود ارتباط من نوع آخر) . وكلما اقتربت ر من الواحد كلما أوحى ذلك بوجود علاقة خطية (موجبة أو سالبة) بين المتغيرين .

ملاحظة (٢)

هناك علاقة مفيدة بين الخطأ المعيارى $ع_{ص.س}$ المعرف بالمعادلة (٨) بالبند (٩ - ٤) في موضوع الانحدار الخطى البسيط ومعامل الارتباط ر وهى :

$$ع^{٢}_{ص.س} = \frac{ن - ١}{ن - ٢} ع^{٢}_{ص} (١ - ر^{٢}) \qquad (٧)$$

ففي المثال (١٠ - ١) نستطيع إيجاد الخطأ المعيارى كالآتي :

$$ع^{٢}_{ص} = \frac{١}{ن-١} [مج ص^{٢} - \frac{(مج ص)^{٢}}{ن}] = \frac{١}{٩} (١٠٨٠ - \frac{١٠٢^{٢}}{١٠}) = ٤,٤$$

إذن $ع^{2}_{ص.س} = \frac{9}{8} \times 4,4 \ (1 - 0,898^{2}) = 0,9583$

إذن $ع_{ص.س} = 0,9789$ تقريبا .

(١٠ – ٥) دلالة معامل الارتباط العزمى :

ليكن ρ هو معامل الارتباط الخطى بين المتغيرين س ، ص في المجتمع ، ر هى معامل الارتباط الناتج من العينة . تحت الافتراضات الثلاثة المذكورة نستطيع اختبار أى فرض عن القيمة الحقيقية للمعامل ρ ووضع حدود ثقة لهذا المعامل .

(أولا) اختبار الفرض ρ = صفر .

لاختبار الفرض الصفرى $ف_{.}$: ρ = صفر (المتغيران غير مرتبطين خطياً)

ضد الفرض الآخر $ف_{1}$: $\rho \neq$ صفر

نستخدم الإحصاءة

$$ت_{.} = \frac{ر\sqrt{ن - 2}}{\sqrt{1 - ر^{2}}} \qquad (8)$$

التي لها توزيع ت بدرجات حرية ن – ٢ بشرط صحة الفرض الصفرى .

مثال (١٠ – ٢) :

في المثال (١٠ – ١) اختبر ما إذا كان المتغيران س ، ص غير مرتبطين خطياً مستخدماً مستوى الدلالة ٠,٠١

الحل :

لدينا ن = ١٠ ، ر = ٠,٨٩٨

$ف_{.} : ρ = ٠ \quad ، \quad ف_{١} : ρ \neq ٠$ (اختبار ذو جانبين)

$$\text{من (٨)} : ت_{ى} = \frac{٠,٨٩٨ \sqrt{١٠ - ٢}}{\sqrt{١ - ٠,٨٩٨^{٢}}} = ٥,٧٧٣$$

$$\text{من الجدول } ت_{٠,٠١[٨]} = ٣,٣٥٥$$

بما أن ٥,٧٧٣ > ٣,٣٥٥ نرفض الفرض الصفرى عند مستوى الدلالة ٠,٠١ ونحكم بوجود ارتباط خطى بين المتغيرين .

ملاحظة (٣)

الإحصاءة (٨) التى تختبر الفرض $ρ = ٠$ ضد الفرض $ρ \neq ٠$ تكافىء الإحصاءة (١٠) بالبند (٩ – ٥) فى موضوع الانحدار الخطى البسيط عند استخدامها لاختبار الفرض $β = ٠$ ضد الفرض $β \neq ٠$ وذلك لأن كلاهما يقيس وجود أو عدم وجود علاقة خطية بين المتغيرين س ، ص . وقد يبدو ذلك غريبا لأن الإحصاءة (٨) تتطلب أن يكون كلا المتغيرين معتدلا بينما الإحصاءة (١٠) تتطلب أن يكون المتغير ص فقط معتدلا . وتزول هذه الغرابة إذا علمنا أن فيشر قد أثبت أنه فى الحالة الخاصة التى يكون فيها $ρ = ٠$ فإن توزيع المعاينة لمعامل الارتباط ر لا يتغير سواء كان المتغير س معتدلا أو غير معتدل طالما كان المتغير ص معتدلا .

ولذلك فإن الإحصاءة (٨) تصلح فقط لاختبار الفرض الصفرى $ρ = ٠$ ولا تصلح لاختبار أى فرض صفرى آخر مثل $ρ = ٠,٥$ أو لايجاد فترات الثقة لمعامل الارتباط $ρ$ فى المجتمع ، وذلك لأنه حين يكون المتغيران مرتبطين ، أى حين $ρ \neq ٠$ يكون توزيع ر ملتويا وبعيدا عن الاعتدالية . فكيف نختبر الفرض $ρ = ك$ حيث $ك \neq ٠$؟

(ثانيا) اختبار الفرض $\rho = \text{ك}$ حيث ك $\neq$ ٠ .

وجد فيشر أنه عندما $\rho \neq$ ٠ فإن التحويل

$$\text{ع} = \frac{1}{2}\,\text{لو}\,\frac{1+\text{ر}}{1-\text{ر}} \quad ، \quad \text{ع}_{\text{م}} = \frac{1}{2}\,\text{لو}\,\frac{1+\rho}{1-\rho} \qquad (٩)$$

يعرف متغيرا عشوائيا ع له توزيع معتدل على وجه التقريب متوسطه $\text{ع}_{\text{م}}$ وتباينه

$$\sigma^2_{\text{ع}} = \frac{1}{\text{ن} - 3} \qquad (١٠)$$

وأن هذا التوزيع يقترب بسرعة من التوزيع المعتدل بزيادة حجم العينة . وبذلك يكون للمتغير

$$\text{ص} = \frac{\text{ع} - \text{ع}_{\text{م}}}{\sigma_{\text{ع}}} \qquad (١١)$$

توزيع معتدل قياسى على وجه التقريب، وبالتالى يجوز تناوله باستخدام جدول المساحات أسفل المنحنى المعتدل القياسى . كما أن الفترة

$$(\text{ع} - \sigma_{\text{ع}} \times \text{مع}_{\frac{\alpha}{2}} \ ، \ \text{ع} + \sigma_{\text{ع}} \times \text{مع}_{\frac{\alpha}{2}}) \qquad (١٢)$$

تكون فترة ثقة بدرجة $(1 - \alpha)$ للبارامتر $\text{ع}_{\text{م}}$ ، حيث $\text{مع}_{\frac{\alpha}{2}}$ هى قيمة المتغير المعتدل القياسي ص التى تحقق المعادلة

$$\text{ل}\,(\,-\text{مع}_{\frac{\alpha}{2}} < \text{ص} < \text{مع}_{\frac{\alpha}{2}}) = 1 - \alpha$$

ومن الفترة (١٢) نستطيع إيجاد فترة الثقة لمعامل الارتباط ρ للمجتمع من التحويل (٩) .

ونتجنب مشقة حساب التحويل (٩) وُضع جدول يحول r إلى ع (بصرف النظر عن الإشارة) هو الجدول (١١) بملحق هذا الكتاب كما وضع جدول يحول ع إلى r هو الجدول (١٢) .

مثال (١٠ – ٣) :

فى عينة عشوائية من ٢٨ زوجا من مجتمع معتدل ذى متغيرين وجد أن معامل الارتباط ٠,٧ . هل هذه القيمة تتمشى مع الفرض القائل أن معامل الارتباط فى المجتمع هو ٠,٥ ؟ أوجد فترة ثقة بدرجة ٩٥٪ لهذا المعامل .

الحل :

الفرض الصفرى : ρ = ٠,٥ الفرض الآخر $\rho \neq$ ٠,٥

نحول كلا من معامل الارتباط فى العينة ومعامل الارتباط المفروض للمجتمع إلى القيمتين المناظرتين لهما باستخدام الجدول (١١) .

r = ٠,٧ تنتج ع = ٠,٨٦٧

ρ = ٠,٥ تنتج ع$_\rho$ = ٠,٥٤٩

الخطأ المعيارى $= \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{25}} = 0.2$

على أساس صحة الفرض الصفرى نجد أن :

$$\therefore \text{ص} = \frac{0.867 - 0.549}{0.2} = 1.59$$

وبما أن ١,٥٩ أقل من القيمة الحرجة ١,٩٦ لا يكون لدينا دليل يدعونا لرفض الفرض الصفرى ، ونحكم بأن معامل الارتباط فى المجتمع يمكن أن يكون ٠,٥ . من الصيغة (١٢) ، نوجد فترة الثقة بدرجة ٩٥٪ للبارامتر ع$_\rho$ كالآتى :

الحد الأدنى للفترة = ٠,٨٦٧ − ٠,٢ × ١,٩٦ = ٠,٤٧٥ = ٠,٤٨ تقريبا

الحد الأعلى للفترة = ٠,٨٦٧ + ٠,٢ × ١,٩٦ = ١,٢٥٩ = ١,٢٦ تقريبا

∴ فترة الثقة للبارامتر ع هى (٠,٤٨ ، ١,٢٦) .

من الجدول (١٢) نجد أن (٠,٤٤٦ ، ٠,٨٥١) هى فترة الثقة بدرجة ٩٥٪ لمعامل الارتباط .

مثال (١٠ − ٤) :

فى عينة من ٩ أزواج من المشاهدات وجد أن معامل الارتباط − ٠,٨٨٩ أوجد فترة ثقة بدرجة ٩٩٪ لمعامل الارتباط فى المجتمع .

الحل :

من الجدول (١١) : ر = − ٠,٨٨٩ وإذن ع = ١,٤١٧ (مع إهمال إشارة ر)

$$\text{الخطأ المعيارى} = \frac{١}{\sqrt{ن - ٣}} = \frac{١}{\sqrt{٦}} = ٠,٤٠٨$$

الحد الأدنى للفترة للبارامتر ع = ١,٤١٧ − ٠,٤٠٨ × ٢,٥٨ = ٠,٣٦٦ .

الحد الأعلى للفترة للبارامتر ع = ١,٤١٧ + ٠,٤٠٨ × ٢,٥٨ = ٢,٤٦٨

∴ الفترة (٠,٣٧ ، ٢,٤٧) هى فترة ثقة بدرجة ٩٩٪ للبارامتر ع . من الجدول (١٢) ومع استرجاع إشارة ر تكون الفترة (−٠,٩٨٦ ، −٠,٣٥٤) هى فترة ثقة بدرجة ٩٩٪ لمعامل الارتباط ρ.

ثالثا) اختبار دلالة الفرق بين معاملى ارتباط عينتين مستقلتين :

نفرض أن لدينا عينتين مستقلتين حجماهما $ن_1$ ، $ن_2$ من الأزواج أعطيتا معاملى ارتباط $ر_1$ ، $ر_2$. نريد أن نختبر ما إذا كان من الممكن اعتبار هاتين العينتين مأخوذتين من نفس المجتمع (أو من مجتمعين لهما نفس معامل الارتباط) ، أى نريد أن نختبر ما إذا كان الفرق بين $ر_1$ ، $ر_2$ ذا دلالة .

نحول القيمتين $ر_1$ ، $ر_2$ إلى القيمتين المناظرتين $ع_1$ ، $ع_2$ باستخدام الجدول (١١) . إذا كانت العينتان مستقلتين ومأخوذتين من نفس المجتمع المعتدل فإن الفرق بين $ع_1$ ، $ع_2$ يكون له توزيع معتدل تباينه يساوى مجموع تباينهما ، أى

$$\sigma^2_{ع_1 - ع_2} = \frac{1}{ن_1 - 3} + \frac{1}{ن_2 - 3} \qquad (13)$$

ويكون الخطأ المعيارى للفرق $|ع_1 - ع_2|$ هو الجذر التربيعى لهذا المقدار .

وعلى ذلك نستطيع اختبار دلالة الفرق بين $ع_1$ ، $ع_2$ (أى بين $ر_1$ ، $ر_2$) بواسطة التوزيع المعتدل .

مثال (١٠ - ٥) :

فى عينة عشوائية من ٥ أبقار وجد أن معامل الارتباط بين الزيادة فى الوزن ومقدار الغذاء المأكول ٠,٨٧ . وفى عينة عشوائية مستقلة حجمها ١٢ بقرة وجد أن معامل الارتباط ٠,٥٦ . فهل هذان المعاملان مختلفان اختلافا جوهريا ؟ استخدم مستوى الدلالة ٥٪ .

الحل :

الفرض الصفرى $ص_1 = ص_2$ والفرض الآخر $ص_1 \neq ص_2$

من الجدول (١١) نجد أن $ر_1 = ٠,٨٧$ ومنها $ع_1 = ١,٣٣٣$

، $ر_2 = ٠,٥٦$ ومنها $ع_2 = ٠,٦٣٣$

تباين الفرق $ع_1 - ع_2$ يساوى مجموع التباينين

$$= \frac{1}{٢} + \frac{1}{٩} = ٠,٦١١$$

∴ الخطأ المعيارى للفرق = $\sqrt{٠,٦١١}$ = ٠,٧٨٢

على أساس صحة الفرض الصفرى $ص_1 = ص_2$ (أى $ع_1 = ع_2$)

$$ص = \frac{(١,٣٣٣ - ٠,٦٣٣) - صفر}{٠,٧٨٢} = ٠,٨٩٥$$

وهذه القيمة تقل عن ١,٩٦ فهى ليست ذات دلالة ولا يسعنا ألا أن نحكم بأن $ص_1 = ص_2$.

(١٠ – ٦) التمييز بين الانحدار والارتباط فى دراسة المشكلات :

كل من الانحدار والارتباط الخطى البسيط يتناول العلاقة الخطية بين متغيرين كميين س ، ص . إلا أن هناك مواقف تستدعى استخدام أسلوب الانحدار ولا يجوز أن تُدرس بأسلوب الارتباط ، كما أن هناك مواقف تستدعى استخدام أسلوب الارتباط ولا يجوز أن تدرس بأسلوب الانحدار . ذلك أن الافتراضات التى يبنى عليها التحليل تختلف فى الانحدار عنها فى الارتباط خاصة فيما يتعلق بنوعية المتغير س ، حيث نفترض فى الانحدار أن س متغير رياضى لا يتأثر بالعوامل العشوائية بينما نفترض فى الارتباط أنه عشوائى . وهذا الاختلاف فى النظر إلى المتغير س يعنى من الناحية العملية اختلافا فى طريقة المعاينة . وعلى ذلك فان اختيار الأسلوب الذى يناسب مشكلة ما – انحدار أو ارتباط – يتوقف على الطريقة التى تتبع فى عملية المعاينة أى فى عملية جمع البيانات .

فلاستخدام أسلوب الانحدار يتطلب الأمر أن يختار الباحث قيما ثابتة من المتغير س يحددها قبل إجراء التجربة ثم يقوم بملاحظة ما يظهر من القيم المناظرة للمتغير

ص عند إجراء التجربة . فمثلا فى دراسة العلاقة بين جرعات دواء مهدىء وقدرة الانسان على حل المشاكل المنطقية يبدأ الباحث بتحديد بضعة جرعات مختلفة التركيب من هذا الدواء (قيم المتغير س) ثم يختار عينة عشوائية من الأفراد يقسمها عشوائيا إلى مجموعات عددها هو عدد الجرعات المختلفة ويعطى لأفراد كل مجموعة واحدة من تلك الجرعات ، ثم يستخدم أحد الاختبارات لقياس القدرة المنطقية لكل فرد فيحصل على قيم للمتغير ص . كذلك فى دراسة العلاقة بين طول نبات ومحتوى النيتروجين فى التربة يبدأ الباحث بتحديد عدة أحواض زراعية تختلف فى محتوى النيتروجين (قيم س) ويزرع النبات فى كل منها ثم يقيس أطوال النباتات بعد فترة من الزمن ليحصل على قيم المتغير ص . فى مثل هاتين الحالتين تكون قيم ص فقط خاضعة للمؤثرات العشوائية ، أما قيم س فتكون قيما ثابتة . وينبغى هنا تناول البيانات بأسلوب الانحدار حيث نقوم بإيجاد معادلة تشير إلى مدى اعتماد المتغير ص على المتغير س ، ويحق لنا أن نصف س بأنه المتغير **المستقل** وأن نصف ص بأنه المتغير **التابع** .

أما فى أسلوب الارتباط فيتطلب الأمر أن يبدأ الباحث باختيار عينة عشوائية من المجتمع ثم يقوم بقياس كل من قيم س ، ص لكل وحدة من وحدات العينة وبذلك تكون جميع القيم السينية والصادية خاضعة للمؤثرات العشوائية إذ يكون لكل منها حرية تامة فى اتخاذ أى قيمة من القيم الممكنة فى المجتمع . فمثلا فى دراسة العلاقة بين طول الذراع وطول الرجل لمجتمع من الأطفال ، يختار الباحث عينة عشوائية من الأطفال ثم يقوم بقياس كل من المتغيرين . كذلك فى دراسة العلاقة بين محتوى الكلسترول فى الدم ووزن الجسم فى مجتمع من المرضى بمرض معين نأخذ عينة عشوائية من هؤلاء المرضى ثم نقيس كلا من هذين المتغيرين . فى مثل هاتين الحالتين تكون قيم كل من المتغيرين س ، ص عشوائية . وينبغى هنا استخدام أسلوب الارتباط حيث نقوم بإيجاد عدد نقدر به الدرجة التى يتغير بها المتغيران معا دون تمييز بين متغير مستقل وآخر تابع .

وفى الحالات التى تتطلب دراستها استخدام أسلوب الانحدار يمكن من الناحية الحسابية إيجاد معامل الارتباط ، ولكن المعامل الناتج يكون مجرد عدد لا معنى له ولا يجوز أن يؤخذ كتقدير للارتباط بين المتغيرين . (وهذا لا يتناقض مع استخدامنا لمربع معامل الارتباط وهو ما سميناه بمعامل التحديد ، فى تحليل الانحدار كما رأينا فى الفصل السابق) .

كذلك ، فى الحالات التى تتطلب دراستها استخدام أسلوب الارتباط يمكن من الناحية الحسابية ايجاد معادلة الانحدار ، ولكن المعادلة الناتجة لا تحقق الهدف منها كمعادلة تنبؤ أو لبيان العلاقة الخطية بين المتغيرين لأن التقديرين أ ، ب يكونان فى هذه الحالة تقديرين متحيزين للبارامترين α ، β. ويتضح ذلك عندما نتذكر أننا فى قياس دقة المقدار

ب فى تقدير البارامتر β نستخدم الخطأ المعيارى $ع_{ص.س}$ / $\sqrt{م م ٢ (س)}$ الذى يعتمد على قيمة م م ٢ (س) = مجـ (س − $\overline{س}$)^٢ − راجع البند (٩ − ٥ − أولا) فإذا لم تكن قيم س ثابتة فإن قيمة هذا الخطأ المعيارى تختلف من عينة إلى أخرى وبالتالى تختلف درجة دقة التقدير من عينة إلى أخرى .

إن التمييز بين الحالات التى تدرس بأسلوب الانحدار وتلك التى تدرس بأسلوب الارتباط هو أمر على درجة كبيرة من الأهمية . وقد لوحظ أن الأمر يختلط فى كثير من الحالات فتعامل مشكلات الانحدار على أنها ارتباط وتعامل مشكلات الارتباط على أنها انحدار ، بل ويصل الأمر أحيانا إلى معاملة المشكلة الواحدة على أنها انحدار وارتباط فى آن واحد ، رغم أن طريقة المعاينة لا تسمح إلا باستخدام واحد فقط من هذين الأسلوبين . ويبدو أن السبب فى هذا اللبس يرجع إلى وجود علاقات رياضية كثيرة بين المعاملات فى هذين النوعين من التحليل ، غير أن العلاقات الرياضية شىء والتحليل الإحصائى شىء آخر .

وجدير بالذكر أنه من الممكن دراسة الانحدار حين يكون كل من المتغيرين س ،

ص عشوائيا ، ولكن ذلك يتطلب استخدام نموذج إحصائى يختلف عن النموذج الذى استخدمناه فى الفصل السابق بحيث يسمح بوجود خطأ عشوائى فى المتغير س . وسوف لا نتعرض لدراسة هذا النموذج خاصة وأنه يفضل دراسة مثل هذه الحالة بأسلوب الارتباط .

(١٠ – ٧) معامل ارتباط الرتب (سبيرمان) :

RANK CORRELATION COEFFICIENT

بالرغم من أن معامل ارتباط بيرسون يستخدم أساسا فى حالة المتغيرات المتصلة إلا أنه يستخدم أيضا فى حالات أخرى . ومن هذه الحالات الحالة التى يتعذر فيها قياس المفردات بالمقياس العددى المعتاد وإنما يمكن قياسها بميزان الترتيب حيث تأخذ كل مفردة ترتيبين س ، ص تعبر الأولى عن ترتيب المفردة بالنسبة للمتغير الأول وتعبر الثانية عن ترتيبها بالنسبة للمتغير الثانى ويكون المطلوب قياس الارتباط بين الترتيبين . ومثال ذلك قيام اثنين من الحكام بترتيب عدد من المتقدمين لشغل وظيفة بحسب أفضليتهم لهذه الوظيفة ثم قياس مدى الاتفاق بين الحكمين . فى هذه الحالة يمكن إثبات أن معامل الارتباط العزمى يأخذ الصيغة البسيطة الآتية التى تعزى إلى سبيرمان ويرمز له بالرمز $\text{ر}_{\text{س}}$ حيث

$$\text{ر}_{\text{س}} = 1 - \frac{6 \sum \text{ف}_{\text{ر}}^2}{\text{ن}(\text{ن}^2 - 1)} \qquad \text{وحيث } \text{ف}_{\text{ر}} = \text{س}_{\text{ر}} - \text{ص}_{\text{ر}} \qquad (14)$$

مثال (١٠ – ٣) :

كانت تراتيب ١٢ طالباً في مادتي الفيزياء س والرياضيات ص كما يلى :

س :	١١	٧	٩	١٢	١	٤	١٠	٣	٨	٥	٢	٦
ص :	١٠	٥	٨	١١	١	٦	١٢	٣	٧	٤	٢	٩

أوجد معامل الارتباط الخطى .

الحل :

بما أن المتغيرين مقاسان بميزان الترتيب فمن الأسهل استخدام معامل ارتباط الرتب كما في الجدول الآتى ، الذى يستخدم الصيغة (١٤) .

الجدول (١٠ – ٢)

إيجاد معامل ارتباط الرتب من بيانات المثال (١٠ – ٣)

س	ص	ف	ف٢
١١	١٠	١	١
٧	٥	٢	٤
٩	٨	١	١
١٢	١١	١	١
١	١	٠	٠
٤	٦	–٢	٤
١٠	١٢	–٢	٤
٣	٣	٠	٠
٨	٧	١	١
٥	٤	١	١
٢	٢	٠	٠
٦	٩	–٣	٩
		صفر	٢٦

بالتعويض في (١٤) نجد أن :

$$ر = ١ - \frac{٦ \times ٢٦}{١٢ \times ١٤٣} = ٠,٩١$$

ونحصل على هذه النتيجة بالضبط إذا استخدمنا معامل ارتباط بيرسون المعرف في (٥) ، مع ملاحظة أنه قد يوجد فرق طفيط يعود إلى عمليات تقريب الأعداد .

(١٠ - ٨) مميزات معامل ارتباط الرتب :

كما سبق القول ، يصلح معامل ارتباط الرتب المعرف في (١٤) لقياس الارتباط بين المتغيرات التى تقاس بميزان الترتيب وهو يصلح أيضاً للمتغيرات التي تقاس بالمقياس العددى المعتاد ولكن يكون من المرغوب فيه تعيين رتب لكل مفردة بدلا من قيمها العددية . ومن أهم مميزات هذا المعامل أنه لا يشترط فرض اعتدالية المتغيرين س ، ص ، بل لا يشترط فرض وجود علاقة خطية بينهما ومن هنا فهو يفضل معامل الارتباط العزمى في الحالة التي لا يتوفر فيها هذان الشرطان ، وقياس الارتباط بواسطة معامل الرتب يعتبر من فصيلة المقاييس غير البارامترية أى التي لا تستلزم وضع فروض معينة على توزيعات المجتمعات .

ومن الأمثلة التي يستخدم فيها هذا المعامل في العلوم البيولوجية قياس الارتباط بين ترتيب انبثاق يرقات مجموعة من الحشرات وترتيب حجومها ، أو بين ترتيب استنبات مجموعة من النباتات وترتيب تزهيرها ، أو لقياس الاتفاق بين اثنين من البيولوجيين في ترتيبهما لمجموعة من الكائنات العضوية من حيث أكثرها شبهاً لصيغة معينة إلى أقلها شبها بها .

مثال (١٠ - ٤) :

في عشرة أنواع من السجائر وجدت المقادير الآتية بالمليجرام من القار والنيكوتين . حول هذه المقادير إلى رتب ثم أوجد معامل ارتباط الرتب لقياس درجة العلاقة بين محتويات القار والنيكوتين .

النـــوع : (١) (٢) (٣) (٤) (٥) (٦) (٧) (٨) (٩) (١٠)
محتـوى القـار : ١٤ ١٧ ٢٨ ١٧ ١٦ ١٣ ٢٤ ٢٥ ١٨ ٣١
محتـوى النيكـوتين : ٠,٩ ١,١ ١,٦ ١,٣ ١,٠ ٠,٨ ١,٥ ١,٤ ١,٢ ٢,٠

الحل :

نرتب كلا من مفردات المجموعتين بحسب الأصغر فالأكبر (أو العكس) كما في الجدول الآتي ، حيث وضعت التراتيب بدلا من القيم المعطاة .

الجدول (١٠ - ٣)

النوع	س	ص	ف	ف^2
(١)	٢	٢	٠	٠
(٢)	٤,٥	٤	٠,٥	٠,٢٥
(٣)	٩	٩	٠	٠
(٤)	٤,٥	٦	١,٥−	٢,٢٥
(٥)	٣	٣	٠	٠
(٦)	١	١	٠	٠
(٧)	٧	٨	١−	١
(٨)	٨	٧	١	١
(٩)	٦	٥	١	١
(١٠)	١٠	١٠	٠	٠
			صفر	٥,٥

بالتعويض فى (١٤) نجد أن $ر_س = ١ - \frac{٦ \times ٥,٥}{١٠ \times ٩٩} = ٠,٩٦٦٧$

ملاحظة (٤) :

إذا كان لمفردتين أو أكثر نفس القيمة العددية ، تعطى لكل منها رتبة تساوى الوسط الحسابي للرتب التي كانت ستأخذها هذه المفردات لو أنها كانت مختلفة القيم ، فمثلا في محتوى القار لدينا عددان متساويان ١٧ ، ١٧ والمفروض أن أحدهما الرابع والآخر الخامس ولذلك أعطى لكل منهما الترتيب $\frac{١}{٢}$ (٤ + ٥) = ٤,٥ .

ملاحظة (٥) :

إذا كانت البيانات أصلا على هيئة أزواج من التراتيب كما فى المثال (١٠ ـــ ٣) فإن معامل الارتباط العزمى يساوى بالضبط معامل ارتباط الرتب . أما إذا كانت البيانات أصلا على هيئة أزواج من القيم العددية كما فى المثال (١٠ ـــ ٤) ، فإن معامل الارتباط العزمى لها لا يساوى معامل ارتباط الرتب الناتج من تحويل هذه القيم إلى رتب .

(١٠ – ٩) دلالة معامل ارتباط الرتب :

إن الإحصاءة التي يعبر عنها معامل ارتباط الرتب $ر_{س}$ المعرف بالصيغة (١٤) توزيعها متماثل حول الصفر . ويمكن إثبات أنه إذا كان المتغيران س ، ص مستقلين فإن هذا التوزيع يقترب من توزيع معتدل وسطه الحسابي صفر وتباينه $\frac{١}{ن - ١}$ حين تقترب ن من اللانهاية . وعلى ذلك عندما يكون حجم العينة كبيراً (ن > ٣٠) نستطيع أن نختبر ما إذا كان هناك ارتباط ذو دلالة بين المتغيرين وذلك بحساب القيمة .

$$ع = \frac{ر_{س} - ٠}{١ / \sqrt{ن - ١}} = ر_{س} \sqrt{ن - ١}$$

ومقارنتها بالقيم الحرجة للتوزيع المعتدل المعيارى ، مع ملاحظة أن الفرض الصفرى هو ص = ٠ والفرض الآخر هو ص ≠ ٠

-راجع المسألة (١ – حـ) من تمارين (٤) .

فمثلا إذا كانت ن = ٥٠ ، r_s = ٠,٤١ ، α = ٠,٠١ فإن :

ع = ٠,٤١ $\sqrt{٤٩}$ = ٢,٨٧

∵ ٢,٨٧ > ٢,٥٨ نرفض الفرض الصفرى عند مستوى الدلالة ٠,٠١ ونحكم بوجود ارتباط بين المتغيرين .

أما إذا كانت ن ⩽ ٣٠ وعلى فرض استقلال المتغيرين أيضاً نستخدم الجدول (١٠) بملحق هذا الكتاب الذى يعطى القيم الحرجة اليمنى عند مستويات الدلالة ٠,٠٥ ، ٠,٠٢٥ ، ٠,٠١ ، ٠,٠٠٥ وحيث ن = ٥ ، ٦ ، ... ، ٣٠ لمقارنتها بالقيم الموجبة لمعامل ارتباط الرتب r_s الناتجة من العينة . وبالنسبة للقيم السالبة لهذا المعامل نستفيد من تماثل التوزيع فى إيجاد القيم الحرجة . ففى المثال (١٠ – ٤) الأخير وجدنا أن r_s = ٠,٩٦٦٧ ، ن = ١٠

الفرض الصفرى ف. : ρ_s = ٠

الفرض الآخر ف١ : ρ_s > ٠ اختبار ذو جانب واحد

من الجدول (١٠) القيمة الحرجة اليمنى عند ن = ١٠ ، α = ٠,٠١ هى ٠,٧٤٥ وحيث أن هذه القيمة أصغر من ٠,٩٦٦٧ نرفض الفرض الصفرى عند مستوى الدلالة ٠,٠١ لصالح الفرض الآخر ونستنتج أن هناك ارتباطاً موجباً بين مقادير القار ومقادير النيكوتين .

ملاحظة (٦) – الارتباط والسببية :

إذا وجدنا أن هناك ارتباطاً بين متغيرين فلا ينبغى أن نسارع بالقول بوجود علاقة سببية بينهما أو بأن أحدهما يحدث نتيجة للآخر . وهناك حالات يكون

فيها هذا القول صحيحاً ويكون التغير في أحد المتغيرين هو فعلا سبب أو أحد أسباب التغير في الآخر كما هو الحال مثلا بين درجة الحرارة وعدد ضربات القلب ، وهناك حالات لا يصح فيها الاستنتاج حيث يكون الارتباط الذى ظهر بين المتغيرين ناشئاً عن وجود متغير ثالث أو أكثر يؤثر فيهما معاً فيحدث هذا الارتباط ، كما هو الحال مثلا بين النجاح في الرياضيات والنجاح في التاريخ حيث يؤثر في كل منهما الذكاء أو المثابرة أو البيئة المنزلية وما إلى ذلك . إن تفسير وجود الارتباط لا يكون بناء على قيمة معامل الارتباط فقط بل على ما لدينا من معلومات عن المتغيرين اللذين ندرسهما .

تمارين (١٠)

في كل من المسائل الخمسة الآتية أوجد معامل الارتباط العزمى (بيرسون) من العينة المعطاة واختبر ما إذا كان المتغيران س ، ص مرتبطين خطياً .

(١) س : ١٥٩ ١٧٩ ١٠٠ ٤٥ ٣٨٤ ٢٣٠ ١٠٠ ٣٢٠ ٨٠ ٢٢٠ ٣٢٠ ٣١٠

ص : ١٤,٤٠ ١٥,٢٠ ١١,٣٠ ٢,٥٠ ٢٢,٧٠ ١٤,٩٠ ١,٤١ ١٥,٨١ ٤,١٩ ١٥,٣٩ ١٧,٣٥ ٩,٥٢

حيث س وزن الخيشوم بالملجرام ، ص وزن الجسم بالجرامات لعينة حجمها ١٢ من نوع من سرطان البحر (أبو جلمبو) .

(٢) س : ١٧ ٢٤ ٢٠ ٢٣ ٢٤ ٢٢ ٢٠ ١٩·٢١ ١٦

ص : ٧ ١٢ ٩ ١٤ ١٢ ١١ ١٠ ٩ ١٠ ٨

حيث س طول نوع معين من نبات البسلة بالمليمترات ، ص الوزن بالمليجرام .

(٣) س : ٥٢ ٤٨ ٥٠ ٥١ ٤٧ ٥١ ٥٢ ٥٢ ٥٣ ٤٨ ٥٠ ٥٢

ص : ٣٦ ٣٤ ٣٤ ٣٤ ٣٥ ٣٥ ٣٦ ٣٦ ٣٧ ٣٤ ٣٤ ٣٧

س : ٥٢ ٥٠ ٥٠ ٤٩ ٤٨ ٤٨ ٤٨ ٥٠ ٥٠ ٤٨ ٥١ ٥٤

ص : ٣٦ ٣٥ ٣٤ ٣٤ ٣٤ ٣٣ ٣٥ ٣٥ ٣٣ ٣٤ ٣٦ ٣٨

حيث س طول جسم طفل حديث الولادة ، ص طول محيط رأسه بالسنتيمترات .

(٤) س : ٥٨ ٥٦ ٥٧ ٥٢ ٦٢ ٥٥ ٦٠ ٥٥ ٦١ ٥٨

ص : ٤١ ٤٣ ٤٢ ٤٤ ٤٤ ٤٠ ٤٢ ٤٥ ٤٥ ٤٤

حيث س ، ص هما أكبر وأصغر قطر لبيض الدجاج بالمليمترات .

(٥) س : ٤,٣ ٤,٥ ٥,٩ ٥,٦ ٦,١ ٥,٢ ٣,٨ ٢,١ ٧,٥

ص : ١٢٦ ١٢١ ١١٦ ١١٨ ١١٤ ١١٨ ١٣٢ ١٤١ ١٠٨

حيث س مقدار المطر في اليوم (٠,٠١ من السنتيمتر) ، ص مقدار مازال من تلوث الهواء نتيجة للمطر (ميكروجرام / متر مكعب) . (لتسهيل الحساب اطرح ١٠٠ من قيم ص) .

(٦) طلب من اثنين من المحكمين ترتيب ١٠ أشخاص متقدمين لوظيفة ما فكانت النتيجة كما يلي . أوجد معامل الارتباط (سبيرمان) واختبر ما إذا كان هناك ارتباط موجب بين المحكمين .

المحكم الأول : ٩ ٤ ٦ ٥ ٣ ١ ١٠ ٨ ٢ ٧

المحكم الثاني : ٨ ٤ ٩ ٦ ١ ٢ ٧ ١٠ ٣ ٥

(٧) الجدول الآتي يبين عدد الساعات (لعينة عشوائية من ١٠ طلاب) التي استذكر فيها هؤلاء الطلاب لاختبار ما وعدد الدرجات التي حصلوا عليها في هذا الاختبار :

عدد ساعات الدراسة : ٨ ٥ ١١ ١٣ ١٠ ٥ ١٨ ١٥ ٢ ٨

درجة الاختبار : ٥٦ ٤٤ ٧٩ ٧٢ ٧٠ ٥٤ ٩٤ ٨٥ ٣٣ ٦٥

حول هذه القيم إلى رتب ثم أوجد معامل ارتباط الرتب (سبيرمان) .

اختبر ما إذا كان هناك ارتباط موجب بين عدد ساعات الدراسة ودرجة الاختبار .

(٨) فى عينتين مستقلتين حجماهما ٢٣ ، ٢٨ من أزواج القيم وجد أن معاملى الارتباط العزمى بين المتغيرين ٠,٥ ، ٠,٨ على الترتيب . اختبر ما إذا كان هذان المعاملان مختلفين اختلافا ذا دلالة .

الفصل الحادى عشر

تحليل التغاير

ANALYSIS OF COVARIANCE

(١١ – ١) التغاير :

لعله من المفيد فى مستهل هذا الفصل أن نذكر القارىء بالمقصود حسابيا بكلمة « التغاير » التى سبق أن وردت فى عدة مناسبات فى هذا الكتاب .

إذا كانت $س_١$ $س_٢$ $س_٣$ .... $س_ن$

، $ص_١$ $ص_٢$ $ص_٣$ ... $ص_ن$

هى ن من أزواج الأعداد فإن تغاير (س ، ص) يعرف بأنه متوسط مجموع حواصل ضرب انحرافات القيم السينية عن متوسطها $\bar{س}$ وانحرافات القيم الصادية عن متوسطها $\bar{ص}$. أى أن :

$$\text{تغاير (س ، ص)} = \frac{1}{ن} \sum_{1}^{ن} (س_ر - \bar{س})(ص_ر - \bar{ص})$$

$$= \frac{1}{ن} \left[\sum_{1}^{ن} س_ر ص_ر - \frac{(\sum س_ر)(\sum ص_ر)}{ن} \right]$$

والمقدار الذى بين القوسين يسمى مجموع حواصل ضرب الانحرافات السينية والانحرافات الصادية عن متوسطيهما أو اختصارا مجموع حواصل الضرب وسنرمز له بالرمز م ص (س ، ص) . وهذا المقدار يمكن أن يكون موجبا أو سالبا أو صفرا .

وكما فى التباين ، إذا كانت أزواج القيم (سر ، صر) هى عينات عشوائية من مجتمع ذى متغيرين وأردنا تقدير التغاير فى هذا المجتمع من التغاير فى العينة فإننا نقسم حواصل الضرب على ن − ١ بدلا من ن وذلك لكى يكون هذا التقدير تقديرا غير متحيز ، أى نكتب :

$$\text{تغاير (س ، ص)} = \frac{1}{ن - 1} \left(\text{مجـ سر صر} - \frac{م\, مَ}{ن} \right) \qquad (2)$$

حيث م مجموع السينات ، مَ مجموع الصادات .

وكما فى التباين أيضا ، ونظرا لأن قيم الانحرافات عن الوسط الحسابى (أو أى قيمة أخرى) لا تتغير بتغير نقطة الأصل فإن قيمة التغاير لا تتغير إذا جمعنا أو طرحنا عددا ثابتا من جميع السينات ، وجمعنا أو طرحنا عددا ثابتا من جميع الصادات .

(١١ − ٢) العلاقة بين تحليل التغاير وتحليل التباين :

فى تحليل التباين بالنموذج ثابت التأثيرات للتجارب ذوات العامل الواحد يكون لدينا متغير كمى ص قسمت قيمه فى عينة ما إلى عدد من المجموعات تناظر مستويات عامل تجريب نوعى مستقل عن المتغير ص . وتهدف الدراسة إلى اختبار دلالة الفروق بين المتوسطات الصادية فى هذه الأقسام لمعرفة مدى تأثير عامل

التجريب عليها . وتتخذ البيانات المشاهدة فى العينة الشكل المبين بالجدول (٨ – ١) بالبند (٨ – ٤) .

إلا أنه فى بعض هذه التجارب يكون المتغير ص واقعا تحت تأثير متغير كمى س يسمى بالمتغير الملازم للمتغير ص Concomitant variable or covariate ، فمثلا قد يكون المتغير ص هو درجات الطلاب فى اختبار رياضيات بعد دراسة مقرر ما فيها ويكون عامل التجريب هو طرق تدريس هذا المقرر . ونظرا لأن استيعاب الطلاب للرياضيات يعتمد على ذكائهم فإن درجاتهم فى الاختبار تكون متأثرة بالذكاء ونقول حينئذ إن المتغير س (نسب ذكاء الطلاب) هو متغير ملازم للمتغير ص . وإذا كان اهتمامنا هو قياس أثر اختلاف مستويات عامل التجريب (طرق التدريس) على المتوسطات الصادية (درجات الرياضيات) عن طريق تحليل التباين فإن دقة هذا القياس تستلزم استبعاد أثر المتغير س (الذكاء) من قيم المتغير ص قبل إجراء هذه العملية .

وفى بعض التجارب يمكن استبعاد هذا الأثر قبل عملية التجريب ، وذلك بأخذ عينات الطلاب التى تختار للتجريب بحيث تكون متكافئة فى الذكاء ، وهنا نستطيع اختبار الفروق بين متوسطات الدرجات بالطريقة المعتادة لتحليل التباين على أساس أن هذه المتوسطات تكون متأثرة فقط بعامل التجريب . غير أن ظروف التجريب قد لا تتيح لنا التحكم فى العينات من حيث جعلها متكافئة فى المتغير الملازم (الذكاء) وهنا يكون لهذا المتغير أثر على المتغير ص (درجات الرياضيات) ولا ينبغى القيام بتحليل التباين إلا بعد استبعاد هذا الأثر . وهذا هو الدور الذى يلعبه تحليل التغاير إذ هو يتألف من شقين أولهما تعديل قيم المتغير ص لاستبعاد أثر المتغير الملازم س وثانيهما تحليل التباين للقيم الصادية المعدلة .

ولكن كيف نصحح القيم الصادية المشاهدة لإزالة أثر هذا المتغير ؟ إذا كان لهذا الأثر وجود فعلىّ فإن انحدار المتغير ص على المتغير س يكون له وجود أيضاً ويكون

تصحيح القيم الصادية المشاهدة $\text{ص}_{\text{رى}}$ عن طريق طرح أثر الانحدار من هذه القيم . وإذا افترضنا أن الانحدار خطى فإن المقدار الذى نطرحه يكون على الصورة ب$(\text{س}_{\text{رى}} - \bar{\bar{\text{س}}})$ حيث ب معامل الانحدار محسوبا من العينة كما سنبين بعد وحيث $\bar{\bar{\text{س}}}$ هى المتوسط العام للقيم السينية فى العينة . أى أننا إذا رمزنا للقيم المعدلة بالرمز $\text{ص}'_{\text{رى}}$ فإن :

$$\text{ص}'_{\text{رى}} = \text{ص}_{\text{رى}} - \text{ب}\,(\text{س}_{\text{رى}} - \bar{\bar{\text{س}}})$$

وإذا أخذنا أى قسم ه على حدة فإن :

$$\text{ص}'_{\text{ر}} = \text{ص}_{\text{ر}} - \text{ب}\,(\text{س}_{\text{ر}} - \bar{\bar{\text{س}}}) \qquad (٢)$$

ويكون علينا بعد ذلك تحليل التباين للقيم الصادية المعدلة .

ومن هنا نرى أن تحليل التغاير لا يعنى تحليل التغاير ذاته كما قد يتبادر إلى الذهن بل يعنى تحليل التباين للمتغير ص بعد عزل أثر المتغير الملازم س مقاسا بمقدار انحدار ص على س .

(١١ – ٣) النموذج الإحصائى :

حين يكون هناك متغير ملازم واحد س لمتغير تابع ص يخضع لعامل واحد من عوامل التجريب له ك من المستويات ، فإن البيانات المشاهدة فى عينة عشوائية تتخذ الشكل المبين بالجدول (١١ – ١) الآتى ، حيث السينات ترمز إلى قيم المتغير الملازم (نسب الذكاء مثلا) والصادات ترمز إلى قيم المتغير التابع (درجات الرياضيات مثلا) ، $\text{ن}_{\text{ى}}$ ترمز إلى عدد القيم فى القسم ه (ه = ١ ، ٢ ، . . . ، ك) ، ن = مجـ $\text{ن}_{\text{ى}}$ = حجم العينة . لاحظ أن هناك قياسين لكل وحدة من وحدات التجريب أحدهما للمتغير التابع وهو $\text{ص}_{\text{رى}}$ والآخر للمتغير الملازم وهو $\text{س}_{\text{رى}}$.

الجدول (١١ – ١)

بيانات تحليل التغاير

الأقسام (المعالجات)									
(١)		(٢)		...	(ى)		...	(ك)	
$س_{١}$	$ص_{١}$	$س_{٢}$	$ص_{٢}$	...	$س_{ى}$	$ص_{ى}$	...	$س_{ك}$	$ص_{ك}$
$س_{١١}$	$ص_{١١}$	$س_{٢١}$	$ص_{٢١}$	...	$س_{١ى}$	$ص_{١ى}$	...	$س_{١ك}$	$ص_{١ك}$
$س_{١٢}$	$ص_{١٢}$	$س_{٢٢}$	$ص_{٢٢}$	...	$س_{٢ى}$	$ص_{٢ى}$	...	$س_{٢ك}$	$ص_{٢ك}$
....		....		....	....		...	...	
...		...		...	...		...	...	
$س_{م١}$	$ص_{م١}$	$س_{م٢}$	$ص_{م٢}$	...	$س_{مى}$	$ص_{مى}$	...	$س_{مك}$	$ص_{مك}$
...		...		...	...		...	...	
...		...		...	...		...	...	
$س_{ن_{١}١}$	$ص_{ن_{١}١}$	$س_{ن_{٢}٢}$	$ص_{ن_{٢}٢}$	...	$س_{ن_{ى}ى}$	$ص_{ن_{ى}ى}$	...	$س_{ن_{ك}ك}$	$ص_{ن_{ك}ك}$
$م_{١}$	$\tilde{م}_{١}$	$م_{٢}$	$\tilde{م}_{٢}$	...	$م_{ى}$	$\tilde{م}_{ى}$	...	$م_{ك}$	$\tilde{م}_{ك}$
$\bar{س}_{١}$	$\bar{ص}_{١}$	$\bar{س}_{٢}$	$\bar{ص}_{٢}$	...	$\bar{س}_{ى}$	$\bar{ص}_{ى}$	....	$\bar{س}_{ك}$	$\bar{ص}_{ك}$

والنموذج الذى نفترضه لتحليل التغاير هو تطوير للنموذج الذى افترضناه فى تحليل التباين بالبند (٨ – ٤ – ١) ، ويتخذ الصيغة الآتية :

$$ص_{مى} = \mu_{ى} + \beta\,(س_{مى} - \bar{س}) + خ_{مى} \qquad (٣)$$

حيث β معامل انحدار ص على س ، م = ١ ، ... ، $ن_{ى}$ ، ى = ١، ٢ ، ...، ك

وهذا النموذج كما نرى يفسح مكانا لأثر انحدار ص على س بالإضافة إلى أثر مستويات عامل التجريب وأثر الخطأ العشوائى .

فى هذا النموذج نفترض الافتراضات المعتادة فى تحليل التباين — راجع البند (٨ — ٤ — ١) — وكذلك الافتراضات المعتادة لتحليل الانحدار — راجع البندين (٩ — ١) ، (٩ — ٥) . وبالإضافة إلى ذلك نفترض فرضا أساسيا فى تحليل التغاير وهو أن العلاقة الخطية بين س ، ص لها نفس معامل الانحدار β فى جميع الأقسام أى يمكن تمثيل هذه العلاقات فى الأقسام المختلفة بخطوط متوازية ، فإذا كانت $\beta_{١}$ ، $\beta_{٢}$ ، ... ، $\beta_{ك}$ هى معاملات الانحدار داخل الأقسام ، فإننا نفترض أن $\beta_{١} = \beta_{٢} = ... = \beta_{ك} = \beta$ وبذلك يمكن ضم البيانات بداخل الأقسام لتعطى تقديرا موحدا للبارامتر β ، وهذا التقدير هو كما نعلم $\frac{\text{٢ مح (س ، ص)}}{\text{٢ ٢ (س)}}$ محسوبا من داخل الأقسام .

إذا كتبنا النموذج (٣) على الصورة

$$\text{ص}_{\text{رو}} - \beta\,(\text{س}_{\text{رو}} - \bar{\text{س}}) = \mu_{\text{ر}} + \text{خ}_{\text{رو}}$$

$$\text{أى} \qquad \text{ص}'_{\text{رو}} \qquad\qquad \mu_{\text{ر}} + \text{خ}_{\text{رو}} \qquad\qquad (٤)$$

يتبين لنا أن تحليل التغاير يتطلب إجراء نوعين من التحليل هما :

١ — تحليل انحدار لتقدير قيمة β ومن ثم تعديل القيم ص$_{\text{رو}}$ المشاهدة إلى القيم ص$'_{\text{رو}}$ لإزالة أثر المتغير الملازم س .

٢ — تحليل التباين للقيم المعدلة ص$'_{\text{رو}}$ لاختبار أثر مستويات التجريب على متوسطاتها . أى أن تحليل التغاير يجمع بين عملية تحليل الانحدار وعملية تتلوها هى عملية تحليل التباين .

(١١ – ٤) خطوات تحليل التغاير :

إن الهدف من تحليل التغاير هو كما سبق القول اختبار دلالة الفروق بين متوسطات المتغير التابع ص بعد تصحيحها لاستبعاد أثر المتغير الملازم س . ومن حقنا أن نتساءل بداية عما إذا كان للمتغير س أثر فعلى على المتغير ص بحيث يستحق عناء عملية الاستبعاد . ولذلك فإن أول ما نهتم به اختبار وجود هذا الأثر أى اختبار وجود علاقة بين س ، ص ، فإذا لم يثبت وجود هذه العلاقة لا يكون هناك ما يدعو لعملية الاستبعاد بل نقوم بتحليل التباين على القيم الصادية المشاهدة دون تعديل . أما إذا ثبت وجودها فينبغى تصحيح هذه القيم قبل إجراء عملية تحليل التباين .

فالخطوة الأولى لتحليل التغاير إذن هى تحليل الانحدار لاختبار أثر المتغير س على المتغير ص ، فإذا ثبتت دلالة الانحدار فإن الخطوة الثانية هى القيام بتحليل التباين للقيم الصادية المعدلة لاختبار أثر عامل التجريب . والمعتاد أن نمهد لهاتين الخطوتين ببعض الحسابات الأولية . وسنوضح ذلك كله بالمثال الآتى .

المثال (١١ – ١) :

فى إحدى التجارب الزراعية الاقتصادية فى قطر ما ، اختيرت ست قرى عشوائيا واختير فى كل منها خمس مزارع عشوائيا . سجلت بالجدول (١١ – ٢) الآتى تكاليف إنتاج محصول الذرة (ص) فى فصل زراعى ما فى كل مزرعة كما سجلت متوسطات المحاصيل (س) التى كانت قد نتجت فى فصول زراعية سابقة فى كل مزرعة . المطلوب استخدام هذه البيانات (أولا) لمعرفة ما إذا كان هناك أثر للمحصول السابق (س) على تكاليف الإنتاج (ص) ، (ثانيا) لمعرفة ما إذا كانت تكاليف الانتاج تختلف من قرية إلى أخرى بعد استبعاد أثر مقدار المحصول السابق إذا كان لهذا الأثر وجود . (لتسهيل الحساب طُرح العدد ١٢ من جميع قيم س والعدد ٣ من جميع قيم ص) .

الجدول (١١ - ٢)

المزارع / القرى	(١) س١	(١) ص١	(٢) س٢	(٢) ص٢	(٣) س٣	(٣) ص٣	(٤) س٤	(٤) ص٤	(٥) س٥	(٥) ص٥	(٦) س٦	(٦) ص٦
(١)	٣	١,٧	٠	٠,٨	٣-	١,٥	٤	٠,٢-	٢-	١,٧	١	٠,٥
(٢)	١-	٢,١	٥	٠,٨-	٠	٠,٦	١		٣-	٢,٣	١-	١
(٣)	٣-	٣,٠	٧	١,٠-	٢-	١,٠	٣	٠,٢-	٤-	٢,٦	٦	٠,٥-
(٤)	٢	١,٣	١-	١,٦	١	٠,٤	٢-	١,٢		١,٠	٢	٠,٧
(٥)	٥	٠,٩	٢	٠,٤	٤-	٢,٥	٠	٠,٢	١-	١,٤	٠	١,٣
المجاميع	٦	٩	١٣	١	٨-	٦	٦	١	١٠-	٩	٨	٣
المتوسطات	١,٢	١,٨	٢,٦	٠,٢	١,٦-	١,٢	١,٢	٠,٢	٢-	١,٨	١,٦	٠,٦

مجـ مجـ س = ١٥ ، $\overline{\text{س}}$ = ٠,٥ ، مجـ مجـ ص = ٢٩ ، $\overline{\text{ص}}$ = ٠,٩٦٦٧

مع ملاحظة أن ن = ٣٠

حسابات تمهيدية :

هذه الحسابات تجرى فى بداية التحليل لخدمة الخطوتين الرئيسيتين المذكورتين ، وهى تتألف من ثلاث مجموعات من الحسابات ترمى المجموعة الأولى إلى إيجاد مجموع المربعات مجـ مجـ (س) للمتغير س وسنرمز له بالرمز أ ثم تحليله بالطريقة المعتادة لتحليل التباين إلى مركبتين إحداهما تعبر عن الاختلاف بين الأقسام وسنرمز له بالرمز أ١ والأخرى تعبر عن الاختلاف المتبقى داخل الأقسام وسنرمز له بالرمز

أ_{2} . وبالمثل بالنسبة لمجموع المربعات م م (ص) للمتغير ص الذى سنرمز له بالرمز ب ولمركبتيه بالرمزين ب_{1} ، ب_{2} ، ثم بالنسبة لمجموع حواصل الضرب م ضـ (س ، ص) الذى سنرمز له بالرمز حـ ولمركبتيه بالرمزين حـ_{1} ، حـ_{2} .

وتعتمد هذه الحسابات على المجاميع الآتية التى يحسن إيجادها مقدما .

$$\text{م}_{1} = \text{مج}\ \text{س}_{\text{ر}1} \ ،\ \text{م}_{2} = \text{مج}\ \text{س}_{\text{ر}2} \ ،\ \ldots \ ،\ \text{م}_{\text{ك}} = \text{مج}\ \text{س}_{\text{رك}} \ ،$$

$$\text{م} = \text{مج}\ \text{مج}\ \text{س} = \overset{\text{ك}}{\underset{1}{\text{مج}}}\ \text{م}_{\text{ى}} \ ،$$

$$\text{مَ}_{1} = \text{مج}\ \text{ص}_{\text{ر}1} \ ،\ \text{مَ}_{2} = \text{مج}\ \text{ص}_{\text{ر}2} \ ،\ \ldots \ ،\ \text{مَ}_{\text{ك}} = \text{مج}\ \text{ص}_{\text{رك}} \ ،$$

$$\text{مَ} = \text{مج}\ \text{مج}\ \text{ص} = \overset{\text{ك}}{\underset{1}{\text{مج}}}\ \text{مَ}_{\text{ى}}$$

$$،\ \text{مج}\ \text{س}^{2} \ ،\ \text{ص}^{2} \ ،\ \text{مج}\ \text{س}\ \text{ص} \ .$$

(١) **تحليل** م م (س)

$\text{أ} = \text{م م (الكلى)} = \text{مج}\ \text{مج}\ \text{س}^{2} - \frac{\text{م}^{2}}{\text{ن}}$ بدرجات حرية ن − ١

$\text{أ}_{1} = \text{م م (بين الأقسام)} = \overset{\text{ك}}{\underset{1}{\text{مج}}}\ \frac{\text{م}_{\text{ى}}^{2}}{\text{ن}_{\text{ى}}} - \frac{\text{م}^{2}}{\text{ن}}$ بدرجات حرية ك − ١

حيث $\text{م}_{\text{ى}}$ مجموع السينات فى القسم ى ، $\text{ن}_{\text{ى}}$ عددها ، ك عدد الأقسام .

$\text{أ}_{2} = \text{م م (داخل الأقسام)} = \text{أ} - \text{أ}_{1}$ بدرجات حرية ن − ك

(٢) تحليل م م (ص)

ب = م م (الكلى) = مج مج ص^{٢} − $\frac{م^{٢}}{ن}$ بدرجات حرية ن − ١

ب$_{١}$ = م م (بين الأقسام) = مج$^{ك}_{١}$ $\frac{م^{٢}_{ى}}{ن_{ى}}$ − $\frac{م^{٢}}{ن}$ بدرجات حرية ك − ١

حيث م$_{ى}$ مجموع الصادات فى القسم ى ، ن$_{ى}$ عددها .

ب$_{٢}$ = م م (داخل الأقسام) = ب − ب$_{١}$ بدرجات حرية ن − ك

(٣) تحليل م مـ (س ، ص)

تسير العمليات الجبرية هنا فى خطوط متوازية مع العمليات الجبرية فى (١) ، (٢) كالآتى :

حـ = م مـ (الكلى) = مج مج س ص − $\frac{م مَ}{ن}$ بدرجات حرية ن − ١

حـ$_{١}$ = م مـ (بين الأقسام) = مج$^{ك}_{١}$ $\frac{مَ_{ى} م_{ى}}{ن_{ى}}$ − $\frac{م مَ}{ن}$ بدرجات حرية ك − ١

حـ$_{٢}$ = م مـ (داخل الأقسام) = حـ − حـ$_{١}$ بدرجات حرية ن − ك

بتطبيق هذه الصيغ على بيانات المثال (١١ − ١) نحصل على مايلى :

(١) تحليل م م (س) :

لدينا م = ٦ + ١٣ + (−٨) + ٦ + (−١٠) + ٨ = ١٥

∴ $\frac{م^{٢}}{ن}$ = $\frac{١٥^{٢}}{٣٠}$ = ٧,٥

ا = م م (الكلى) = [٣٢ + (-١)٢ + ... + (٢)٢ + ٠] - ٧,٥
= ٢٥١,٥ بدرجات حرية ٢٩

ا١ = مم (بين الأقسام) = ١/٥ [٦٢ + ١٣٢ + (-٨)٢ + ٦٢ +
(-١٠)٢ + ٨٢] - ٧,٥

= ٨٦,٣ بدرجات حرية ٥

ا٢ = مم (داخل الأقسام) = ٢٥١,٥ - ٨٦,٣ = ١٦٥,٢بدرجات حرية ٢٤

(٢) تحليل م م (ص) :

لدينا مَ = ٩ + ١ + ٦ + ١ + ٩ + ٣ = ٢٩

∴ مَ٢/ن = ٢٩٢/٣٠ = ٢٨,٠٣

ب = مم (الكلى) = [١,٧٢ + ٢,١٢ + ... + (٠,٧)٢ + ١,٣٢]
- ٢٨,٠٣
= ٢٨,٩٣ بدرجات حرية ٢٩

ب١ = مم (بين الأقسام) = ١/٥[٩٢ + ١٢ + ٦٢ + ١٢ + ٩٢ + ٣٢]
-٢٨,٠٣

= ١٣,٧٧ بدرجات حرية ٥

ب٢ = مم (داخل الأقسام) = ٢٨,٩٣ - ١٣,٧٧ = ١٥,١٦
بدرجات حرية ٢٤

(٣) تحليل م صه (س ، ص) :

لدينا م = ١٥ ، مَ = ٢٩ وإذن $\frac{م مَ}{ن} = \frac{١٥ \times ٢٩}{٣٠} = ١٤,٥$

حـ = م صه (الكلى) $= [٣ \times ١,٧ + (-١ \times ٢,١) + ... + ٢ \times$
$٠,٧ + ٠] - ١٤,٥$
$= - ٥٤,٨ - ١٤,٥ = -٦٩,٣$

بدرجات حرية ٢٩

حـ١ = م صه (بين الأقسام) $= \frac{١}{٥}[٦ \times ٩ + ١٣ \times ١ .. + ٨ \times ٣]$
$- ١٤,٥$

$= - ٨,٢ - ١٤,٥ = -٢٢,٧$ بدرجات حرية ٥

حـ٢ = م صه (داخل الأقسام) $= -٦٩,٣ - (-٢٢,٧) = ٤٦,٦$

بدرجات حرية ٢٤

يحسن تسجيل هذه المجموعات الثلاث من القيم فى جدول واحد كالجدول (١١ - ٣) الآتى ليسهل الرجوع إليها .

الجدول (١١ - ٣)

تحليل مجموعى المربعات ومجموع حواصل الضرب

مصدر التباين	درجات الحرية	م م (س)	م م (ص)	م صه (س ، ص)
بين الأقسام (القرى)	ك - ١ = ٥	أ١ = ٨٦,٣	ب١ = ١٣,٧٧	حـ١ = -٢٢,٧
داخل الأقسام (البواقى) خطأ التجريب	ن - ك = ٢٤	أ٢ = ١٦٥,٢	ب٢ = ١٥,١٦	حـ٢ = -٤٦,٦
الكلى	ن - ١ = ٢٩	أ = ٢٥١,٥	ب = ٢٨,٩٣	حـ = -٦٩,٣

الخطوة الأولى : اختبار تأثير المتغير س على المتغير ص :

تهدف هذه الخطوة إلى اختبار وجود علاقة خطية بين المتغير الملازم س والمتغير التابع ص ، ويتأتى هذا كما نعلم إما بتقدير معامل الانحدار الموحد β من داخل أقسام العينة ثم اختبار دلالة هذا التقدير باختبار ت ، أو بتقدير التباين المفسر (الناشيء عن الانحدار) واختبار دلالته باختبار ف بالصورة

$$ف = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}} \qquad \text{بدرجتى حرية ١ ، ن − ك − ١} \quad (٥)$$

على أن تحسب المقادير اللازمة لهذا الاختبار من **داخل الأقسام** أى من السطر الخاص بالبواقى بالجدول (١١ – ٣) .

فى المثال ، وباستخدام الصيغة (١٧) بالبند (٩ – ٧) نجد ما يلى :

$$\text{الاختلاف المفسر} = \frac{[\text{م مـ (س ، ص)}]^{٢}}{\text{م م (س)}} = \frac{جـ_{٢}^{٢}}{أ_{٢}}$$

$$= \frac{(-٤٦,٦)^{٢}}{١٦٥,٢} = ١٣,١٤٥ \qquad \text{بدرجة حرية واحدة}$$

$\therefore$ الاختلاف غير المفسر = الاختلاف فى ص – الاختلاف المفسر

$$= ب_{٢} - \frac{جـ_{٢}^{٢}}{أ_{٢}} = ١٥,١٦ - ١٣,١٤٥$$

$$= ٢,٠١٥ \qquad \text{بدرجات حرية ٢٣}$$

$$\text{إذن} \qquad ف_{ى} = \frac{١٣,١٤٥ \div ١}{٢,٠١٥ \div ٢٣} = \frac{١٣,١٤٥}{٠,٠٨٧٦} = ١٥٠,٠٦^{**}$$

وهذه القيمة تزيد كثيرا عن القيمة الحرجة ف$_{٠٫٠١[١ ، ٢٣]}$ فهى ذات دلالة عالية ونحكم بوجود علاقة خطية سالبة بين المتغير الملازم س والمتغير التابع ص ، وهذا يعنى أن مقدار المحصول السابق يؤثر فيما تدفعه القرى من تكاليف لإنتاج المحصول الجديد .

ملاحظات :

(١) استخدمنا نفس الصيغ التى تستخدم فى حالة وجود عينة من أزواج القيم (س ، ص) موضوعة فى قسم واحد (ك = ١) لأن هذه الصيغ تظل صالحة عند وجود أكثر من قسم (ك > ١) ، بشرط تعديل درجات الحرية وفقا لذلك أى وضع ن − ك − ١ بدلا من ن − ٢ للاختلاف غير المفسر .

(٢) من الجدول (١١ − ٣) نستطيع إيجاد ثلاثة تقديرات لتباين المتغير ص بعد التصحيح للانحدار ، غير أن التقدير الذى نحصل عليه من داخل الأقسام وهو تباين البواقى حول خط الانحدار يكون هو التقدير غير المتحيز الذى يعكس الخطأ العشوائى فى هذا التحليل ، ونختبر دلالة أى تقدير آخر بالمقارنة به . ولذلك فإن حساب قيمة ف من الصيغة (٥) ينبغى أن يكون من داخل الأقسام كما سبق القول .

الخطوة الثانية : تعديل المتوسطات واختبار دلالة الفروق بينها :

كما سبق القول ، لا تتخذ هذه الخطوة إلا إذا ظهر من الخطوة السابقة وجود أثر فعلى للمتغير الملازم س على المتغير ص . وترمى هذه الخطوة إلى تعديل القيم الصادية المشاهدة بحيث تكون القيم المعدلة مستقلة عن أثر المتغير س . ولما كان هذا التعديل من شأنه تعديل تباين القيم الصادية فإن تحليل التباين لاختبار دلالة الفروق بين المتوسطات المعدلة يكون باستخدام اختبار ف بالصيغة المعتادة مع وضع التباينات المعدلة بدلا من التباينات الأصلية ، أى باستخدام الصيغة

$$ف = \frac{\text{التباين المعدل بين الأقسام}}{\text{التباين المعدل داخل الأقسام}} \qquad (٦)$$

بدرجتى حرية ك − ١ ، ن − ك − ١

فى المثال (١١ − ١) وجدنا أن المتغير الملازم س يؤثر فى المتغير التابع ص وعلى ذلك فإن دقة البحث تقتضى أن نخلص المتغير ص من أثر المتغير س قبل إجراء تحليل التباين ، أى تقتضى استخدام اختبار ف بالصيغة (٦) . ولقد سبق لنا حساب التباين المعدل داخل الأقسام وهو التباين غير المفسر ٢,٠١٥ ÷ ٢٣ = ٠,٠٨٧٦ وهو كما سبق القول تقدير غير متحيز لتباين المتغير ص بعد تعديله .

ومن الصيغة (١٩) بالبند (٩ − ٧) نذكر أن هذا التباين هو مربع الخطأ المعيارى $ع_{ص.س}$ للتقدير من معادلة الانحدار ، أى أن :

$ع^{٢}_{ص.س}$ = التباين غير المفسر (مأخوذا من داخل الأقسام) (٧)
= ٠,٠٨٧٦ فى هذا المثال.وبدرجات حرية ٢٣

أما التباين المعدل بين الأقسام فنوجده بطريقة غير مباشرة بطرح الاختلاف المعدل داخل الأقسام من الاختلاف الكلى المعدل ثم القسمة على درجات الحرية . أى أن :
الاختلاف المعدل بين الأقسام = الاختلاف الكلى المعدل − الاختلاف المعدل داخل الاقسام (٨)

$$= (ب - \frac{ح^{٢}}{أ}) - (ب_{٢} - \frac{ح^{٢}_{٢}}{أ_{٢}}) \quad (٨َ)$$

وفى هذا المثال نجد أن

$$\text{الاختلاف المعدل بين الأقسام} = (٢٨,٩٣ - \frac{(-٦٩,٣)^{٢}}{٢٥١,٥}) - ٢,٠١٥$$

= (٢٨,٩٣ − ١٩,٠٩٥) − ٢,٠١٥
= ٧,٨٢٠ بدرجات حرية ٥

من الصيغة (٦) : $ف_ى = \frac{٧,٨٢٠ \div ٥}{٢,٠١٥ \div ٢٣} = \frac{١,٥٦٤}{٠,٠٨٧٦}$

$$= ١٧,٨٥^{**}$$

وهذه القيمة ذات دلالة عالية لأن $ف_{٠,٠١[٥ ، ٢٣]}$ لا تزيد عن ٣,٩ مما يشير إلى أن متوسط تكاليف الإنتاج – بعد استبعاد أثر مقادير المحاصيل السابقة – ليست متساوية فى القرى الست . ويمكننا إذا أردنا أن نضع هذه النتائج فى الجدول (١١ – ٤) الآتى .

الجدول (١١ – ٤)

اختبار ف للتغاير

مصدر الاختلاف	الاختلاف المشاهد ب	الانحدار حـ٢/أ	الاختلاف المعدل ب – حـ٢/أ	د ح	ط ٢	ف
بين الأقسام (القرى)	–	–	٧,٨٢٠	٥	١,٥٦٤	١٧,٨٥**
داخل الأقسام	١٥,١٦	١٣,١٤٥	٢,٠١٥	٢٣	٠,٠٨٧٦	
الكلى	٢٨,٩٣	١٩,٠٩٥	٩,٨٣٥	٢٨		

ملاحظة :

إذا أهملنا المتغير الملازم س وقمنا بتحليل التباين لقيم ص المشاهدة دون تعديل نجد من بيانات الجدول (١١ – ٣) أن

$$ف_ى = \frac{\text{التباين بين الأقسام}}{\text{التباين داخل الأقسام}} = \frac{١٣,٧٧ \div ٥}{١٥,١٦ \div ٢٤} = \frac{٢,٧٥٤}{٠,٦٣١٧}$$

$$= ٤,٣٦^{**} \qquad \text{بدرجتى حرية ٥ ، ٢٤}$$

وبالرغم من أن هذه القيمة ذات دلالة عالية أيضا لأن $ف_{٠,٠١[٥ ، ٢٤]} = ٣,٩٠$ إلا أنها تقل كثيرا عن القيمة ١٧,٨٥ . هذا مع تذكر أن تحليل التباين يجزىء الاختلاف الكلى فى ص إلى المركبتين بين وداخل الأقسام ، أما تحليل التغاير فهو يجزىء البواقى من التحليل الشامل للانحدار .

(١١ – ٥) المقارنة بين المتوسطات المعدلة :

بعد تحليل التباين للقيم الصادية المعدلة يبقى أن نتدارس كيفية إجراء المقارنات بين متوسطات هذه القيم فى الأقسام المختلفة ، وهذا يتطلب أن نحسب هذه المتوسطات ثم نختبر دلالة الفروق بينها .

(ا) حساب المتوسطات المعدلة :

من الصيغة (٢) بالبند (١١ – ٢) نجد أن المتوسط المعدل $\overline{ص'_ى}$ فى أى قسم يساوى المتوسط المشاهد المناظر $\overline{ص_ى}$ مطروحا منه أثر الانحدار الخطى . أى أن

$$\overline{ص'_ى} = \overline{ص_ى} - ب (\overline{س_ى} - \overline{س}) \qquad (٩)$$

حيث ب معامل انحدار ص على س ويحسب من **داخل الأقسام** بالصيغة المعتادة للانحدار الخطى البسيط كالآتى – انظر الصيغة (٨) بالبند (٩ – ٣) .

$$ب = \frac{م\ مـ (س ، ص)}{م\ م (س)} = \frac{حـ_٢}{ا_٢} \qquad (١٠)$$

ففى المثال : $ب = \frac{-٤٦,٦}{١٦٥,٢} = -٠,٢٨٢$

من جدول البيانات (١١ − ٢) والصيغة (٩) نجد أن المتوسطات المعدلة هى :

$\overline{ص_١}$ = ١,٨ + ٠,٢٨٢ (١,٢ − ٠,٥) = ١,٩٩٧٤

$\overline{ص_٢}$ = ٠,٢ + ٠,٢٨٢ (٢,٦ − ٠,٥) = ٠,٧٩٢٢

$\overline{ص_٣}$ = ١,٢ + ٠,٢٨٢ (−١,٦− ٠,٥) = ٠,٦٠٧٨

وبالمثل نجد أن $\overline{ص_٤}$ = ٠,٣٩٧٤ ، $\overline{ص_٥}$ = ١,٠٩٥ ، $\overline{ص_٦}$ = ٠,٩١٠٢

يلاحظ أن المجموع الكلى للقيم الصادية المعدلة يساوى المجموع الكلى للقيم الصادية المشاهدة وكلاهما يساوى ٢٩ . كما يلاحظ أن بعض المتوسطات صغرت والبعض الآخر كبر ، وأن الفرق بين أى متوسطين معدلين يقل عن الفرق بين المتوسطين المشاهدين المناظرين ، مما يشير إلى أن المتغير الملازم كان له تأثير فعلى على هذه المتوسطات .

(ب) اختبار دلالة الفروق بين أزواج المتوسطات المعدلة :

يلزمنا هنا إيجاد تقدير للخطأ المعيارى $ع_ع$ للفرق بين متوسطين معدلين ، بحيث ندخل فى اعتبارنا الخطأ المعيارى $ع_{ص.س}$ للتقدير من معادلة الانحدار . والصيغة التى نشتق منها الخطأ المعيارى المطلوب هى :

$$ع^٢_ع = ع^٢_{ص.س} \left(١ + \frac{أ_١ \div (ك - ١)}{أ_٢}\right) \text{ بدرجات حرية } ن - ك - ١ \quad (١١)$$

حيث $أ_١$ = مجموع المربعات لقيم س محسوبا من بين أقسام المعالجة

، $أ_٢$ = مجموع المربعات لقيم س محسوبا من داخل أقسام المعالجة

ففى المثال ، وباستخدام الجدول (١١ − ٣) نجد أن

$$ع^٢_ع = ٠,٠٨٧٦ \left(١ + \frac{٨٦,٣ \div ٥}{١٦٥,٢}\right) = ٠,٠٩٦٧٥ \quad \text{بدرجات حرية ٢٣}$$

ونكون الآن مستعدين لإجراء المقارنات بين أزواج المتوسطات الصادية المعدلة باستخدام الأسلوب المبين بالبند (٨ – ٥) أو أى أسلوب مكافىء .

فمثلا للمقارنة بين المتوسطين المعدلين $\overline{ص_١}$ = ١,٩٩٧٤ ، $\overline{ص_٢}$ = ٠,٧٩٢٢ ومع ملاحظة أن مجموعى القيم الصادية فى القسمين (١) ، (٢) هما (بالضرب فى خمسة) ٩,٩٨٧ ، ٣,٩٦١ يمكن أن نستخدم الأسلوب الآتى :

$$م م (١ \text{ ضد } ٢) = \frac{٩,٩٨٧^٢ + ٣,٩٦١^٢}{٥} - \frac{١٣,٩٤٥^٢}{١٠} = ٣,٦٢٩٣$$

$$\therefore ف_ى = \frac{٣,٦٢٩٣ \div ١}{٠,٠٩٦٧٥} = ٣٧,٥١٢^{**}$$ بدرجتى حرية ١ ، ٢٣ .

وهذه القيمة ذات دلالة عالية مما يجعلنا نرفض الفرض الصفرى عن تساوى متوسطى التكاليف فى القريتين (١) ، (٢) .

كما يمكننا هنا استخدام اختبار ت بالصيغة

$$ت = \frac{\overline{ص_١} - \overline{ص_٢}}{ع_ع \sqrt{\frac{١}{ن_١} + \frac{١}{ن_٢}}}$$ بدرجات حرية ٢٣

فنجد أن $ت_ى$ = ٦,١٢٧ وهى أيضا ذات دلالة عالية ، مع ملاحظة أن مربع هذه القيمة وهو ٣٧,٥٤ يساوى قيمة $ف_ى$ = ٣٧,٥١٢ والفرق يرجع إلى أخطاء التقريب .

تمرين (١١ – ١)

فى جزء من تجربة عن تأثير نوعين من الأدوية فى علاج مرض الجذام كان الهدف مقارنة ثلاثة أنواع من المعالجة : $د_١$ ، $د_٢$ دواءان من صنف المضادات الحيوية ،

د$_{٣}$ دواء داخلى يتخذ كمراقبة . اختير عشرة مرضى للتجريب وحددت على كل منهم ٦ مواقع من الجسم يتراكم فيها ميكروب الجذام وقيست غزارة الجذام باختبارات معملية فى هذه المواقع قبل بداية التجربة (س) ثم بعد عدة شهور من العلاج (ص) ودونت النتائج فى الجدول الآتى .

الجدول (١١ – ٥)

المرضى	الدواء					
	د$_{١}$		د$_{٢}$		د$_{٣}$	
	س$_{١}$	ص$_{١}$	س$_{٢}$	ص$_{٢}$	س$_{٣}$	ص$_{٣}$
(١)	١١	٦	٦	٠	١٦	١٣
(٢)	٨	٠	٦	٢	١٣	١٠
(٣)	٥	٢	٧	٣	١١	١٨
(٤)	١٤	٨	٨	١	٩	٥
(٥)	١٩	١١	١٨	١٨	٢١	٢٣
(٦)	٦	٤	٨	٤	١٦	١٢
(٧)	١٠	١٣	١٩	١٤	١٢	٥
(٨)	٦	١	٨	٩	١٢	١٦
(٩)	١١	٨	٥	١	٧	١
(١٠)	٣	٠	١٥	٩	١٢	٢٠
المجاميع	٩٣	٥٣	١٠٠	٦١	١٢٩	١٢٣
المتوسطات	٩,٣	٥,٣	١٠	٦,١	١٢,٩	١٢,٣

أجر تحليل التغاير لاختبار دلالة الفروق بين متوسطات غزارة الجذام فى المستويات الثلاثة للمعالجة ، بعد استبعاد أثر الاختلاف فى غزارة المرض قبل بدء المعالجة .

(١١ – ٦) تحليل التغاير فى التجارب ذوات العاملين ومتغير ملازم واحد :

النموذج الإحصائى هنا هو امتداد طبيعى للنموذج الإحصائى (٣) للتجارب ذوات العامل الواحد ، وهو يأخذ الصيغة الآتية :

$$\text{ص}_{\text{رو}} = \mu_{\text{و}} + \mu_{\text{ر}} + \beta(\text{س}_{\text{رو}} - \bar{\text{س}}) + \text{خ}_{\text{رو}} \qquad (١٢)$$

ويسير التحليل بنفس الأسلوب مع بعض الإضافات والتعديلات التى يقتضيها وجود عامل تجريبى ثان .

مثال (١١ – ٢) :

فى تجربة لمقارنة مقادير المحاصيل الناتجة من ٦ أنواع من الذرة أخذت عينة عشوائية من ٢٤ حوضا زراعيا وزرعت عليها أنواع الذرة عشوائيا – أربعة أحواض لكل نوع – وقد نتجت البيانات المدونة بالجدول (١١ – ٦) الآتى حيث ص تعبر عن مقدار المحصول الناتج من الحوض ، س تعبر عن درجة خصوبة الحوض مقدرة بعدد النباتات التى كانت قائمة به . إن مقدار المحصول يكون واقعا تحت تأثير عاملين تجريبيين هما : الأحواض (٤ مستويات) وأنواع الذرة (٦ مستويات) ، غير أن المطلوب هنا اختبار ما إذا كان معدل المحصول يختلف باختلاف نوع الذرة (بعد استبعاد أثر الخصوبة) .

الحسابات التمهيدية :

نقوم بتحليل كل من م م (س) ، م م (ص) ، م حـ (س ، ص) إلى ثلاث مركبات : بين الأعمدة وبين الصفوف والخطأ . أى أن هناك خطوة إضافية واحدة فى كل تحليل .

الجدول (١١ – ٦)

أنواع الذرة	الأحواض (١) س١	ص١	(٢) س٢	ص٢	(٣) س٣	ص٣	(٤) س٤	ص٤	المجاميع س	ص	المتوسطات $\bar{س}$	$\bar{ص}$
ا	٢٨	٢٠٢	٢٢	١٦٥	٢٧	١٩١	١٩	١٣٤	٩٦	٦٩٢	٢٤	١٧٣
ب	٢٣	١٤٥	٢٦	٢٠١	٢٨	٢٠٣	٢٤	١٨٠	١٠١	٧٢٩	٢٥,٥	١٨٢,٢٥
جـ	٢٧	١٨٨	٢٤	١٨٥	٢٧	١٨٥	٢٨	٢٢٠	١٠٦	٧٧٨	٢٦,٥	١٩٤,٥
د	٢٤	٢١٠	٢٨	٢٣١	٣٠	٢٣٨	٣٠	٢٦١	١١٢	٩٣١	٢٨	٢٣٢,٧٥
هـ	٣٠	٢٠٢	٢٦	١٧٨	٢٦	١٩٨	٢٩	٢٢٦	١١١	٨٠٤	٢٧,٧٥	٢٠١
و	٣٠	٢٢٨	٢٥	٢٢١	٢٧	٢٠٧	٢٤	٢٠٤	١٠٦	٨٦٠	٢٦,٥	٢١٥
المجاميع	١٦٢	١١٦٦	١٥١	١١٨١	١٦٥	١٢٢٢	١٥٤	١٢٢٥	٦٣٢	٤٧٩٤	٢٦,٣٣	١٩٩,٧٥

مجـ مجـ $س^2$ = ١٦٨٢٤ ، مجـ مجـ $ص^2$ = ٩٧٦٢٨٠ ، مجـ مجـ س ص = ١٦٧٧٢٧

تحليل م م (س) :

$$ا = م\ م\ (\text{الكلي}) = مجـ\ مجـ\ س^2 - \frac{م^2}{ن}$$

$$= ١٦٨٢٤ - \frac{٦٣٢^2}{٢٤} = ١٨١,٣٣٤$$ بدرجات حرية ن – ١ = ٢٣

$$ا_1 = م\ م\ (\text{بين الأعمدة}) = \frac{1}{٦}(١٦٢^2 + ١٥١^2 + ١٦٥^2 + ١٥٤^2) - \frac{٦٣٢^2}{٢٤}$$

= ٢١,٦٦٧ بدرجات حرية ك – ١ = ٣

$أ_1$ = مم (بين الصفوف) = $\frac{1}{4}$ (٩٦^٢ + ١٠١^٢ + ... + ١٠٦^٢) − $\frac{٦٣٢^٢}{٢٤}$

= ٤٥,٨٣٤ بدرجات حرية هـ − ١ = ٥

∴ $أ_2$ = مم (الخطأ) = ١٨١,٣٣٤ − (٢١,٦٦٧ + ٤٥,٨٣٤)

= ١١٣,٨٣٣ بدرجات حرية (ك − ١) (هـ − ١) = ١٥

وبنفس الطريقة نحلل كلا من م م (ص) و م حـ (س ، ص) ونضع النتائج فى الجدول الآتى حيث ك ترمز إلى عدد الأعمدة ، هـ ترمز إلى عدد الصفوف .

الجدول (١١ − ٧)

تحليل مجموعى المربعات ومجموع حواصل الضرب

مصدر التباين	درجات الحرية	م م (س)	م م (ص)	م حـ (س ، ص)
بين الأعمدة (الأحواض)	ك − ١ = ٣	$أ_1$ = ٢١,٦٧	$ب_1$ = ٤٣٦,١٧	$حـ_1$ = ٨,٥٠
بين الصفوف (الأنواع)	هـ − ١ = ٥	$أ'$ = ٤٥,٨٣	$ب'$ = ٩٤٩٠,٠٠	$حـ'$ = ٥٥٩,٢٥
خطأ التجريب	(ك−١)(هـ−١)=١٥	$أ_2$ = ١١٣,٨٣	$ب_2$ = ٨٧٥٢,٣٣	$حـ_2$ = ٩١٧,٢٥
أنواع + خطأ	٢٠	١٥٩,٦٦	١٨٢٤٢,٣٣	١٤٧٦,٥
الكلى	ن − ١ = ٢٣	أ = ١٨١,٣٣	ب = ١٨٦٧٨,٥٠	حـ = ١٤٨٥,٠٠

وستتبين أهمية السطر الإضافى (أنواع + خطأ) عند تكوين النسبة ف لاختبار المتوسطات المعدلة .

الخطوة الأولى : اختبار تأثير المتغير اللازم س على المتغير ص :

أى اختبار وجود علاقة خطية بين المتغيرين س ، ص ويتأتى هذا عن طريق الصيغة (٥) لاختبار ف مع ملاحظة درجات الحرية .

ف = التباين المفسر / التباين غير المفسر بدرجتى حرية ١ ، ن − ك − هـ (١٣)

على أن يحسب البسط والمقام من **السطر الخاص بخطأ التجريب** . فى المثال نجد ما يأتى :

الاختلاف المفسر = $\frac{ح_٢^٢}{أ_٢}$ = $\frac{٩١٧,٢٥^٢}{١١٣,٨٣}$ = ٧٣٩١,٢٦٣٨

بدرجة حرية واحدة

∴ الاختلاف غير المفسر = $ب_٢ - \frac{ح_٢^٢}{أ_٢}$ = ٨٧٥٢,٣٣ − ٧٣٩١,٢٦٣٨

= ١٣٦١,٠٦٦٢ بدرجات حرية ١٤

∴ $ف_ى$ = $\frac{٧٣٩١,٢٦٣٨ \div ١}{١٣٦١,٠٦٦٢ \div ١٤}$ = $\frac{٧٣٩١,٢٦٣٨}{٩٧,٢١٩٠}$

= $٧٦,٠٢٧^{**}$ بدرجتى حرية ١ ، ١٤

وهذه القيمة ذات دلالة عالية مما يدل على وجود تأثير كبير لخصوبة الأرض على معدل المحصول . ولذلك ينبغى استبعاد هذا الأثر قبل مقارنة متوسطات المحاصيل تحت تأثير أنواع الذرة .

الخطوة الثانية : اختبار دلالة الفروق بين المتوسطات المعدلة :

نستخدم الصيغة (٦) لاختبار ف مع ملاحظة درجات الحرية .

$$\text{ف} = \frac{\text{التباين المعدل (بين الأنواع)}}{\text{التباين المعدل للخطأ}} \quad \text{بدرجتى حرية ك}-1 \text{ ، ن}-\text{ك}-\text{هـ} \quad (١٤)$$

ولقد سبق أن حسبنا فى الخطوة الأولى التباين المعدل للخطأ فهو التباين غير المفسر
١٣٦١,٠٦٦٢ ÷ ١٤ = ٩٧,٢١٩

أما التباين المعدل بين الأنواع فيحسب كالآتى :

التباين المعدل بين الأنواع = [الاختلاف المعدل (أنواع + خطأ) – الاختلاف المعدل للخطأ] مقسوما على درجات الحرية (١٥)

من السطر الإضافى بالجدول (١١ – ٧) نجد ما يلى

$$\text{الاختلاف المعدل (أنواع + خطأ)} = ١٨٢٤٢,٣٣ - \frac{{١٤٧٦,٥}^2}{١٥٩,٦٦}$$

= ٤٥٨٧,٩٨٩ بدرجات حرية ١٩

ولكن الاختلاف المعدل للخطأ = ١٣٦١,٠٦٦٢ بدرجات حرية ١٤

∴ الاختلاف المعدل بين الأنواع = ٤٥٨٧,٩٨٩ – ١٣٦١,٠٦٦٢

= ٣٢٢٦,٩٢٢٨ بدرجات حرية ٥

∴ التباين المعدل بين الأنواع = ٣٢٢٦,٩٢٢٨ ÷ ٥ = ٦٤٥,٣٨٥

من الصيغة (١٤) ينتج أن :

$$\text{ف}_{ى} = \frac{٦٤٥,٣٨٥}{٩٧,٢١٩} = ٦,٦٣٨^{**} \quad \text{بدرجتى حرية ٥ ، ١٤}$$

وهذه القيمة ذات دلالة عالية لأنها أكبر من $ف_{٠,٠١ [٥ ، ١٤]}$ التى لا تزيد عن ٥,٠٦ مما يشير إلى أن متوسطات المحاصيل (بعد استبعاد أثر الخصوبة) ليست جميعها متساوية عند الأنواع المختلفة من الذرة .

ملاحظة :

يمكن بنفس الطريقة اختبار ما إذا كان معدل المحصول يختلف باختلاف الأحواض (الأعمدة) .

(١١ – ٧) المقارنة بين أزواج المتوسطات المعدلة :

نستخدم نفس الأسلوب المقدم بالبند (١١ – ٥ – ب) ، فتحسب المتوسطات المعدلة من الصيغة (٩) وهى :

$$\overline{ص_و}' = \overline{ص_و} - ب\,(\overline{س_و} - \overline{س})$$

حيث معامل الانحدار $ب = \frac{م\,ص\,(س ، ص)}{م\,م\,(س)} = \frac{ح_٢}{أ_٢} = \frac{٩١٧,٢٥}{١١٣,٨٢} = ٨,٠٥٨٨$

ومن هذه القيمة نتوصل إلى المتوسطات المعدلة الآتية :

$\overline{ص_١}' = ١٩١,٨$ ، $\overline{ص_٢}' = ١٩١,٠$ ، $\overline{ص_٣}' = ١٩٣,١$

$\overline{ص_٤}' = ٢١٩,٣$ ، $\overline{ص_٥}' = ١٨٩,٦$ ، $\overline{ص_٦}' = ٢١٣,٦$

أما الخطأ المعيارى $ع_خ$ للفرق بين متوسطين معدلين فيحسب من الصيغة

$$ع^٢_خ = ع^٢_{ص.س}\left[١ + \frac{أ_١ \div (هـ - ١)}{أ_٢}\right] \text{بدرجات حرية } ن - ك - هـ \quad (١٦)$$

حيث $أ_١$ = مجموع المربعات لقيم س محسوبا من بين الأنواع

، $أ_٢$ = مجموع المربعات لقيم س محسوبا من خطأ التجريب .

$$ع^٢_ح = ٩٧,٢١٩ \; (١ + \frac{٤٥,٨٣ \div ٥}{١١٣,٨٣})$$

$$= ١٠٥,٠٤٨ \qquad \text{بدرجات حرية ١٤}$$

وبهذه القيمة نكون قادرين على مقارنة ما نريد من أزواج المتوسطات الصادية المعدلة كالمعتاد .

يمتد تحليل التغاير إلى المواقف الأكثر تعقيدا . وكمثال لذلك التجارب ذوات العامل الواحد التى تشتمل على متغيرين ملازمين $س_١$ ، $س_٢$ يرتبطان خطيا بالمتغير الرئيسى ص ، حيث يأخذ النموذج الإحصائى الشكل الآتى :

$$ص_{ري} = \mu_ر + \beta_١ (س_{١ري} - \overline{س}_١) + \beta_٢ (س_{٢ري} - \overline{س}_٢) + خ_{ري}$$

ويحتاج الأمر هنا إلى تحليل الانحدار المتعدد قبل القيام بتحليل التباين .

تمرين (١١ - ٢)

بالجدول الآتى مقادير المحاصيل (ص) لنبات فول الصويا فى الفدان ونسبة الإصابة (س) لساق النبات بمرض معين . استخدمت أربعة خطوط أ ، ب ، حـ ، ء للمقارنة وزرع بكل منها ٤ نباتات . المطلوب (١) توضيح أن هناك علاقة خطية سالبة بين الإصابة بالمرض ومقدار المحصول (٢) توضيح أن معدل المحاصيل يختلف من خط إلى آخر بعد استبعاد أثر الإصابة (٣) ايجاد المتوسطات المعدلة للمحاصيل والمقارنة بين كل زوج منها (٤) توضح أنه إذا لم يُستبعد أثر الإصابة فإن معدل المحاصيل لا يختلف من خط إلى آخر .

الجدول (١١ - ٨)

القطاعات	الخطوط أ		ب		جـ		د		المجاميع	
	$س_١$	$ص_١$	$س_٢$	$ص_٢$	$س_٣$	$ص_٣$	$س_٤$	$ص_٤$	س	ص
(١)	١٩,٣	٢١,٣	١٠,١	٢٨,٣	٤,٣	٢٦,٧	١٤,٠	٢٥,١	٤٧,٧	١٠١,٤
(٢)	٢٩,٢	١٩,٧	٣٤,٧	٢٠,٧	٤٨,٢	١٤,٧	٣٠,٢	٢٠,١	١٤٢,٣	٧٥,٢
(٣)	١,٠	٢٨,٧	١٤,٠	٢٦,٠	٦,٣	٢٩,٠	٧,٢	٢٤,٩	٢٨,٥	١٠٨,٦
(٤)	٦,٤	٢٧,٣	٥,٦	٣٤,١	٦,٧	٢٩,٠	٨,٩	٢٩,٨	٢٧,٥	١٢٠,٢
المجاميع	٥٥,٩	٩٧,٠	٦٤,٤	١٠٩,١	٦٥,٥	٩٩,٤	٦٠,٣	٩٩,٩	٢٤٦,١	٤٠٥,٤

مجـ مجـ $س^٢$ = ٦٤٦٥,٩٢٥٦، مجـ مجـ $ص^٢$ = ١٠٦٣١,٩٢٢ ، مجـ مجـ س ص = ٥٣٥٢,٠٥٨٧

الفصل الثانى عشر

الانحدار والارتباط الخطى المتعدد

MULTIPLE LINEAR REGRESSION AND CORRELATION

سنعتمد فى هذا الفصل على الحاسب الإلكترونى فى إجراء الحسابات اللازمة لحل مشكلات الانحدار والارتباط تخففا من الأعباء الحسابية الضخمة التى يتطلبها التحليل . إلا أن هذا لا يعفينا بل يوجب علينا الإحاطة بالأسس والافتراضات والمفاهيم التى يبنى عليها التحليل تحسبا من أخطاء ومزالق التطبيق .

أولا : الانحدار الخطى المتعدد

(١٢ - ١) الانحدار الخطى المتعدد كامتداد للانحدار الخطى البسيط :

فى تناولنا للانحدار الخطى البسيط فى الفصل التاسع من هذا الكتاب افترضنا أن لدينا متغيرا عشوائيا ص نرغب فى التنبؤ بقيمه عن طريق قيم متغير غير عشوائى س على أساس أن العلاقة بين هذين المتغيرين هى علاقة خطية :

ص = $\alpha + \beta$ س حيث α ، β بارامتران مجهولان ،

وعلى أساس أنه عند أى قيمة ثابتة س يكون للمتغير ص توزيع معتدل متوسطه $\alpha + \beta$ س يتوقف على قيمة س ، وتباينه عدد ثابت σ^2 لا يتوقف على س .

إلا أنه فى كثير من التطبيقات يرى الباحث أن استخدام متغير واحد س لا يصلح للتنبؤ بقيم المتغير ص بدقة كافية ، ويأمل أن يحصل على تنبؤات أكثر دقة إذا استخدم عدة متغيرات $س_١$ ، $س_٢$ ، . . . ، $س_ك$ يشعر بخبرته فى ميدان عمله أن لها دورا فى عملية التنبؤ . وإذا تبنينا افتراضات مماثلة لتلك التى تبنيناها فى الانحدار الخطى البسيط من أن هذه المتغيرات غير عشوائية وترتبط بالمتغير العشوائى ص بعلاقة خطية :

$$ص = \alpha + \beta_١ س_١ + \beta_٢ س_٢ + \ldots + \beta_ك س_ك \qquad (١)$$

حيث α ، $\beta_١$ ، $\beta_٢$ ، . . . ، $\beta_ك$ بارامترات مجهولة ، ومن أنه عند أى مجموعة ثابتة $س_١$ ، $س_٢$ ، ... ، $س_ك$ يكون للمتغير ص توزيع معتدل ذو تباين ثابت $\sigma^٢$ ، فإن الانحدار يسمى فى هذه الحال بالانحدار الخطى المتعدد أو اختصارا بالانحدار الخطى كما تسمى البارامترات $\beta_١$ ، $\beta_٢$ ، . . . ، $\beta_ك$ بمعاملات الانحدار الجزئية partial regression coefficients ، وتكون دراسة هذا الانحدار مجرد تعميم للمفاهيم والأساليب والصيغ والاختبارات المستخدمة فى الانحدار الخطى البسيط ولا تختلف عنها إلا فى درجة المشقة فى تناول البيانات ، كما تختلف بطبيعة الحال فى درجات الحرية التى تعتمد على عدد الثوابت فى معادلة الانحدار التى تقدر من العينة . وإذا كان القارىء قد اطمأن لمعرفته بأسس وأساليب الانحدار الخطى البسيط فسوف لا يجد صعوبة تذكر فى تناول الانحدار الخطى المتعدد كما يتبين من الفقرات الآتية .

(أ) ايجاد معادلة الانحدار :

فى الانحدار الخطى البسيط نقوم بتجربة نحصل منها على عينة من ن من المشاهدات ($س_ر$ ، $ص_ر$) تأخذ الشكل الآتى :

س١ س٢ ... سن

ص١ ص٢ ... صن

وتستخدم هذه البيانات فى تقدير البارامترين المجهولين α ، β بقيمتين ا ، ب توطئة لكتابة معادلة الانحدار :

ص̂ = ا + ب س

التى تمثل « أحسن خط » يلائم تلك المشاهدات من وجهة نظر مبدأ المربعات الصغرى ، الذى يستلزم كما نعلم تحديد القيمتين أ،ب اللتين تجعلان الدالة :

مج (صر − ص̂ر)٢ = مج (صر − α − β سر)٢

نهاية صغرى . وتحقيق هذا الفرض يستلزم بدوره إجراء عملية التفاضل الجزئى بالنسبة إلى كل من α ، β ومساواة كل من الناتجين بالصفر للحصول على معادلتين خطيتين فى هذين المجهولين تسميان بالمعادلتين المعتادتين وهما :

مج ص = ن α + β مج س

مج س ص = α مج س + β مج س٢

وحل هاتين المعادلتين معا يعطينا القيمتين ا ، ب المطلوبتين .

وهذه الخطوات هى بذاتها الخطوات التى تتخذ عند تناول الانحدار الخطى المتعدد . إذ نقوم بتجربة نحصل منها على عينة من ن من المشاهدات (س١ر ، س٢ر ، .... ، سكر ، صر) حيث ر = ١ ، ٢ ، ... ، ن تأخذ الشكل المبين بالجدول (١٢ − ١) الآتى :

الجدول (١٢ - ١)

بيانات الانحدار الخطى المتعدد

المشاهدات	ص	$س_١$	$س_٢$	....	$س_ك$
(١)	$ص_١$	$س_{١١}$	$س_{٢١}$	...	$س_{١ك}$
(٢)	$ص_٢$	$س_{١٢}$	$س_{٢٢}$	...	$س_{٢ك}$
...	...	...	...	...	
(ر)	$ص_ر$	$س_{١ر}$	$س_{٢ر}$	...	$س_{رك}$
...	...	...	...	...	...
(ن)	$ص_ن$	$س_{١ن}$	س	...	$س_{نك}$

ولتقدير البارامترات المجهولة α ، $\beta_١$ ، $\beta_٢$ ، ... ، $\beta_ك$ من هذه البيانات نستخدم مبدأ المربعات الصغرى لايجاد القيم ا ، $ب_١$ ، $ب_٢$ ، .... ، $ب_ك$ توطئة لكتابة معادلة الانحدار :

$$\hat{ص} = ا + ب_١ س_١ + ب_٢ س_٢ + ... + ب_ك س_ك \qquad (٢)$$

وتحقيق هذا المبدأ يستلزم أن تؤخذ القيم أ،$ب_١$ ، ....، $ب_ك$ بحيث تكون الدالة

$$\sum (ص_ر - \hat{ص}_ر)^٢ = \sum (ص_ر - \alpha - \beta_١ س_{ر١} - \beta_٢ س_{ر٢} - ... - \beta_ك س_{رك})^٢$$

نهاية صغرى ، وهذا يستلزم بدوره إجراء عملية التفاضل الجزئى بالنسبة إلى كل من α ، $\beta_١$ ، $\beta_٢$ ، .. ، $\beta_ك$ ومساواة النواتج بالصفر للحصول على ك + ١ من المعادلات الخطيه فى هذه المجاهيل تسمى بالمعادلات المعتادة . وحل هذه المعادلات معا يعطينا التقديرات المطلوبه ا ، $ب_١$ ، $ب_٢$ ، .. ، $ب_ك$ وعددها ك + ١ .

ولنا أن نتصور هنا مدى المشقة التى نلقاها فى حساب المجاميع ومجاميع المربعات ومجاميع حواصل الضرب المطلوبة لوضعها فى هذه المعادلات ثم حل المعادلات ذاتها ، خاصة إذا كان عدد المتغيرات السينية أكبر من اثنين . وحتى فى حالة وجود متغيرين $س_١$ ، $س_٢$ حيث نستهدف الوصول إلى معادلة الانحدار

$$\hat{ص} = ا + ب_١ س_١ + ب_٢ س_٢$$

يكون علينا حل المعادلات المعتادة الآتية :

$$مج ص = ن \alpha + \beta_١ مج س_١ + \beta_٢ مج س_٢$$

$$مج س_١ ص = \alpha مج س_١ + \beta_١ مج س_١^٢ + \beta_٢ مج س_١ س_٢ \quad (٣)$$

$$مج س_٢ ص = \alpha مج س_٢ + \beta_١ مج س_١ س_٢ + \beta_٢ مج س_٢^٢$$

وهذا الحل يحتاج أولا إلى حساب ثمانية مجاميع هى مج ص ، مج $س_١$ ، مج $س_٢$ ، مج $س_١^٢$ ، مج $س_٢^٢$ ، مج $س_١ س_٢$ ، مج $س_١$ ص ، مج $س_٢$ ص مما يحتاج إلى كثير من الجهد والوقت . ولذلك نلجأ إلى الحاسب الالكترونى ليقوم عنا بكل هذه العمليات ويعطينا التقديرات المطلوبة بكل سرعة ودقة .

(ب) إيجاد الخطأ المعيارى لتقدير ص من خط الانحدار :

فى الانحدار الخطى البسيط استخدمنا مقياسا رأينا أهميته البالغة فى عمليات الاستنتاج الإحصائى هو الخطأ المعيارى للتقدير الذى رمزنا له بالرمز $ع_{ص.س}$ وعرفناه فى البند (٩ — ٤) بأنه الانحراف المعيارى للقيم المعيارية المشاهدة $ص_ر$ حول القيم $\hat{ص}_ر$ المقدرة من معادلة الانحدار أى عرفناه كالآتى :

$$ع^٢_{ص.س} = \frac{١}{ن - ٢} مج (ص_ر - \hat{ص}_ر)^٢ \qquad \text{بدرجات حرية ن — ٢}$$

وحين حللنا الاختلاف الكلى فى القيم الصادية وهو مج $(ص_ر - \bar{ص})^{٢}$ إلى مركبتين تعبر الأولى وهى مج $(\hat{ص}_ر - \bar{ص})^{٢}$ عن الاختلاف الناشىء عن الانحدار (الاختلاف المفسر) وتعبر الثانية وهى مج $(ص_ر - \hat{ص}_ر)^{٢}$ عن الاختلاف المتبقى الناشىء عن الانحراف عن خط الانحدار (الاختلاف غير المفسر) أمكننا كتابة مربع الخطأ المعيارى للتقدير كالآتى :

$$ع^{٢}_{ص.س} = \frac{١}{ن - ٢} \text{ مج } (ص_ر - \hat{ص}_ر)^{٢} = \frac{١}{ن - ٢} \times \text{ الاختلاف غير المفسر}$$

بنفس المنطق نستخدم نفس التعريف فى حالة الانحدار الخطى المتعدد مع مراعاة درجات الحرية ، فنرمز للخطأ المعيارى لتقدير ص من خط الانحدار بالرمز

$ع_{ص.س_١ س_٢ ... س_ك}$ حيث :

$$ع^{٢}_{ص.س_١ س_٢ ... س_ك} = \frac{١}{ن - (ك + ١)} \times \text{ الاختلاف غير المفسر} \qquad (٤)$$

حيث ن حجم العينة ، ك عدد المتغيرات السينية ومع ملاحظة أن ك + ١ هو عدد الثوابت فى معادلة الانحدار التى تقدر من العينة .

ملاحظة (١) :

ليس من المناسب هنا تقديم الصيغة العامة للخطأ المعيارى للتقدير ويكفينا أن نفهم المعنى الذى يتضمنه ، أما حساب قيمته فنتركه للحاسب الالكترونى . إلا أنه حين يكون هناك متغيران تنبؤيان اثنان فقط (ك = ٢) فيمكن إثبات أن

$$ع^{٢}_{ص.س_١ س_٢} = \frac{١}{ن - ٣} (\text{مج} ص^{٢} - أ \text{ مج} ص - ب_١ \text{ مج} س_١ ص - ب_٢ \text{ مج} س_٢ ص) \quad (٥)$$

وهذه الصيغة هى امتداد للصيغة (١٠) بالبند (٩ – ٤) التى تتناول الحالة التى يكون لدينا فيها متغير تنبؤى واحد .

هذا مع ملاحظة أن الصيغة العامة للخطأ المعيارى تكتب بأسلوب رياضى يتطلب معرفته دراسة مسبقة لموضوع المصفوفات ، والواقع أن الدراسة النظامية للانحدار المتعدد تعتمد برمتها على هذا الأسلوب غير أننا فى التطبيق العملى لا نحتاج إليه .

(حـ) اختبار دلالة الانحدار ككل :

فى الانحدار الخطى البسيط اهتممنا باختبار دلالة الانحدار أى باختبار وجود علاقة خطية بين المتغير التابع ص والمتغير المستقل س وهذا يكافىء اختبار الفرض الصفرى $\beta = ٠$ ضد الفرض $\beta \neq ٠$ واستخدمنا لهذا الغرض اختبار ت بالصيغة (١١) بالبند (٩ – ٥ – أولا) مع وضع $\beta =$. أو اختبار ف بالصيغة (٢١) بالبند (٩ – ٧) أو بالصيغة المكافئة (٢٤) ، بشرط توفر شروط الانحدار .

كذلك يهمنا فى حالة الانحدار الخطى المتعدد اختبار وجود علاقة خطية بين المتغير التابع ص والمتغيرات المستقلة $س_{١}$ ، $س_{٢}$ ، . . . ، $س_{ك}$ وهذا الاختبار يكافىء اختبار الفرض الصفرى $\beta_{١} = \beta_{٢} = . . . = \beta_{ك} = ٠$ وفى هذه الحالة نستخدم أى من الصيغتين (٢١) أو (٢٤) المذكورتين مع مراعاة استخدام درجات الحرية المناسبة . وهاتان الصيغتان هما :

$$ف = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}} \quad \text{بدرجتى حرية ك ، ن – ك – ١} \qquad (٦)$$

$$\text{أو } ف = \frac{ر^{٢} \div ك}{(١ - ر^{٢}) \div (ن - ك - ١)} \quad \text{بدرجتى حرية ك ، ن – ك – ١} \qquad (٧)$$

حيث ن حجم العينة ، ك عدد المتغيرات السينية ، $ر^2$ هو معامل التحديد الذى يعبر كالمعتاد عن نسبة الاختلاف فى المتغير ص التى يفسرها خط الانحدار وسنرمز له هنا بالرمز $ر^2_{ص(١٢...ك)}$ حيث

$$ر^2_{ص(١٢...ك)} = \frac{\text{الاختلاف المفسر}}{\text{الاختلاف الكلى}} \qquad (٨)$$

(د) اختبار دلالة معاملات الانحدار الجزئية :

فى الانحدار الخطى البسيط يكون لدينا معامل انحدار واحد β ، ويمكن أن نختبر ما إذا كان هذا المعامل يأخذ قيمة ما نفترضها بواسطة الإحصاءة (١١) المقدمة بالبند (٩ – ٥ – أولا) وهى

$$ت = \frac{\mathbf{B} - \beta}{ع_{ص.س} / \sqrt{م م (س)}} = \frac{\mathbf{B} - \beta}{ع (ب)} \qquad (٩)$$

التى يكون لها توزيع ت بدرجات حرية ن – ٢ بشرط تحقق افتراضات الانحدار . ويلاحظ أن مقام هذا الكسر وهو ع (ب) = $ع_{ص.س} / \sqrt{م م (س)}$ هو تقدير للانحراف المعيارى للمتغير العشوائى $\mathbf{B}$ الذى تعتبر ب إحدى قيمه المشاهدة – وحيث ب تقدير معامل الانحدار β الذى نجده من العينة . والمتغير $\mathbf{B}$ هو متغير معتدل وسطه الحسابى β وانحرافه المعيارى σ (ب) يقدر من العينة بالمقدار ع (ب) .

أما فى الانحدار الخطى المتعدد فلدينا ك من معاملات الانحدار الجزئية β_1 ، β_2 ، ... ، $\beta_ك$ (بالإضافة إلى البارامتر α) ويهمنا أن نختبر ما إذا كانت هذه المعاملات تأخذ قيما معينة . وبنفس منطق البند (٩ – ٥) نعتبر أن القيمة $س_ى$

($ى = ١ ، ٢ ، ... ، ك$) التى نحصل عليها من عينة لتقدير البارامتر $\beta_{ى}$ ، هى إحدى القيم المشاهدة من متغير عشوائى $B_{ى}$ ذى توزيع معتدل متوسطه $\beta_{ى}$ وانحرافه المعيارى σ ($ب_{ى}$) نقدره بقيمة سنرمز لها بالرمز ع ($ب_{ى}$) وبالتالى يكون للإحصاءة

$$ت_{ى} = \frac{B_{ى} - \beta_{ى}}{ع\ (ب_{ى})} \qquad ى = ١ ، ٢ ، ... ، ك \qquad (١٠)$$

توزيع ت بدرجات حرية $ن - (ك + ١)$ بشرط تحقق افتراضات الانحدار . وهذه الإحصاءة تصلح لاختبار أن يأخذ أى معامل $\beta_{ى}$ أى قيمة نفترضها ، وبصفة خاصة لاختبار الفرض $\beta_{ى} = ٠$ ، لأهمية ذلك فى بحث مدى مساهمة المتغير المناظر $س_{ى}$ فى التنبؤ من معادلة الانحدار .

ملاحظة (٢) :

ليس من المناسب هنا أيضا تقديم الصيغة العامة التى نوجد منها التقدير ع ($ب_{ى}$) للانحراف المعيارى للمتغير $B_{ى}$ وسنعتمد فى ذلك أيضا على الحاسب الالكترونى ويكفينا أن نحيط بالدور الذى يقوم به . إلا أنه حين يكون هناك متغيران تنبؤيان اثنان فقط ($ك = ٢$) فيمكن إثبات أن هذا التقدير يحسب من الصيغة الآتية :

$$ع\ (ب_{ى}) = \frac{ع_{ص.س_{١}س_{٢}}}{\sqrt{(١ - ر^{٢}_{٢١})\ .\ م\ م\ (س_{ى})}} \qquad ى = ١ ، ٢ \qquad (١١)$$

حيث $ع^{٢}_{ص.س_{١}س_{٢}} = \frac{١}{ن - ٣} \times$ الاختلاف غير المفسر

هو مربع الخطأ المعيارى للتقدير ويمكن حسابه من الصيغة (٥) ، وحيث $ر^{٢}_{٢١}$ هو معامل التحديد للمتغيرين $س_{١}$ ، $س_{٢}$ ويمكن حسابه من الصيغة :

$$ر^{2}_{21} = \frac{[\text{م ص } (س_1 ، س_2)]^{2}}{\text{م م } (س_1) \cdot \text{م م } (س_2)} \qquad (12)$$

كما أن م م ($س_1$) هو مجموع المربعات للمتغير س ، م م ($س_2$) هو مجموع المربعات للمتغير $س_2$.

(١٢ - ٢) استخدام الحاسب الالكترونى :

يتضح من البند السابق أن حل المشكلات العملية فى تحليل الانحدار يتطلب القيام بعمليات حسابية طويلة سواء فى إيجاد المجاميع أو حل المعادلات أو فى استخراج تقديرات للأخطاء المعيارية أو فى استخراج قيم ت أو قيم ف اللازمة للاستنتاج الإحصائى ، مما يستغرق جهدا شاقا ووقتا طويلا ، إضافة إلى صعوبة تلافى أخطاء هذه الحسابات إذا أجريت يدويا . ولذلك فإن معظم المشكلات التطبيقية تستعين بالحاسبات الالكترونية التى تنوب عنا فى استخراج ما نريد بكل سرعة ودقة وتترك لنا فقط مهمة تفسير النتائج واتخاذ القرارات بناء على ما قدمته من معلومات .

واستخدام هذه الحاسبات لا يستلزم أن يكون الباحث قادرا على تشغيلها بنفسه بل يكفيه أن يتصل بأحد مراكز الحاسبات وتقديم مصفوفة البيانات التى حصل عليها من التجربة مع تحديد مجموعة الأعداد التى تخص كلا من المتغير الصادى والمتغيرات السينية ، وتحديد ما يريد إيجاده من معلومات .

وهناك كثير من الحاسبات تختزن برامج إحصائية تقوم بكافة أنواع التحليل ، ومن هذه البرامج ما يلى :

STATPACK, MINITAB, COSAP AND SPSS

ومثل هذه البرامج من شأنها توفير الجهد والوقت مع ضمان الدقة والأمان ولذلك يعتمد هذا الفصل فى حساباته وتحليلاته على الحاسبات الالكترونية التى أصبحت

اليوم فى متناول الجميع . وسنوضح ذلك بالمثال الآتى الذى يتناول حالة متغيرين تنبؤيين $س_{١}$ ، $س_{٢}$. على أن الأسلوب المستخدم يمتد إلى الحالات التى تتناول أكثر من متغيرين دون أى تعديل .

مثال (١٢ – ١) :

أرادت إحدى الشركات التجارية الكبيرة أن تعد طريقة للتنبؤ بعدد الوحدات التى تباع فى الشهر فى أى فرع من فروعها العشرة ، على افتراض أن الوحدات المباعة فى أى فرع تتأثر بعاملين هما (١) عدد البائعين فى الفرع ، (٢) مقدار ما يصرفه الفرع شهريا على الإعلان . وتحقيقا لهذا الغرض جمعت بيانات من الفروع ودونت فى الجدول (١٢ – ٣) الآتى .

الجدول (١٢ – ٣)

الفرع	عدد الوحدات المباعة ص	عدد العاملين $س_{١}$	تكاليف الاعلان $س_{٢}$
(١)	٢٥	٥	١٠
(٢)	٢٠	٢	١١
(٣)	٣٠	٦	١٢
(٤)	٢٥	٤	١٣
(٥)	٢٥	٣	١٤
(٦)	٣٢	٦	١٥
(٧)	٢٥	٤	١٢
(٨)	٢١	٣	١١
(٩)	٢٠	٢	١٠
(١٠)	٢٧	٥	١٢

المطلوب إيجاد معادلة الانحدار الخطى للمتغير ص على المتغيرين س١ ، س٢ واختبار دلالة الانحدار ككل ودلالة كل من معاملى الانحدار .

الحل :

للتوضيح سنقوم بحل هذه المسألة مرتين : باستخدام الحاسب ثم بدونه .

(أولا) الحل باستخدام الحاسب :

أدخلت البيانات المدونة بالجدول (١٢ – ٣) فى حاسب الكترونى فأخرج المعلومات المدونة بالجدول (١٢ – ٤) الآتى ، مع ملاحظة أن عدد المتغيرات التنبؤية ك = ٢ وأن حجم العينة ن = ١٠ .

الجدول (١٢ – ٤)

مخرجات الحاسب الالكترونى لبيانات المثال (١٢ – ١)

(1)	Regress.of	Y	Number of Units Sold		
(2)	on	X_1	Number of Salespeople		
(3)		X2	Amount of Advertising Expenditure		
(4)	Variable Name	Regress. Coeff	S.E. of Coeff.	t	D.F.
(5)	Constant	6.05263			
(6)	X1	2.10526	.17708	11.88891	7
(7)	X2	.87719	.16125	5.42652	7

(8) Coefficient of Determination (R ^ 2) = .974659

(9) Estimated Standard Error of Estimate = .722013

Analyis of Variance for Regression

(10) Source of Variation	SS	D.F.	MS	F
(11) Regression	140.351	2	70.1754	134.615
(12) Residual	3.64912	7	.521303	

(ا) معادلة الانحدار :

بالتأمل فى الجدول (١٢ – ٤) نرى أن الحاسب قد قام بإيجاد التقديرات المطلوبة للبارامترات α ، β_1 ، β_2 وهى:

ا = ٦,٠٠٥٢٦٣ ، $ب_١$ = ٢,١٠٥٢٦ ، $ب_٢$ = ٠,٨٧٧١٩

وهذه القيم الثلاث هى تلك المدونة بالصفوف ٥ ، ٦ ، ٧ من العمود الثانى بالجدول . وإذن معادلة الانحدار هى – التقريب إلى ٣ خانات عشرية :

$\hat{ص}$ = ٦,٠٠٥٣ + ٢,١٠٥ $س_١$ + ٠,٨٧٧ $س_٢$

فمثلا إذا كان عدد البائعين فى الفرع $س_١$ = ٤ والتكاليف الشهرية $س_٢$ = ١٢ فإننا نتوقع أن يكون عدد الوحدات المباعة حوالى ٢٤ وحدة .

(ب) اختبار دلالة الانحدار الخطى :

من الجدول (١٢ – ٤) نجد أن الحاسب قد حسب لنا جدول التباين فى السطور الثلاثة الأخيرة ، فقد أعطانا كلا من :

الاختلاف المفسر (الانحدار) = ١٤٠,٣٥١ بدرجات حرية ك = ٢

الاختلاف غير المفسر (الانحراف عن خط الانحدار) = ٣,٦٤٩١٢ بدرجات حرية ن – ك – ١ = ٧ بل أعطانا أيضا نسبة التباين $ف_ى$ مستخدما الصيغة (٦) وهى

$ف_ى$ = التباين المفسر / التباين غير المفسر = ٧٠,١٧٥٤ / ٠,٥٢١٣٠٣ = ١٣٤,٦١٥

الاستنتاج :

لما كانت القيمة ١٣٤,٦١٥ تزيد كثيرا عن القيمة الحرجة $ف_{٠,٠١[٢ ، ٧]}$ = ٩,٥٥

نرفض الفرض الصفرى أن $\beta_1 = \beta_2 = 0$ عند مستوى عالى من الدلالة بمعنى أن هناك علاقة خطية بين عدد الوحدات المباعة من ناحية وعدد البائعين وتكاليف الإعلان من ناحية أخرى .

كما أن الحاسب قد أعطانا قيمة معامل التحديد بين المتغير ص والمتغيرين $س_1$ ، $س_2$ وهو $ر^2_{ص(1،2)}$ = ٠,٩٧٤٦٥٩ وهذا يعنى أن المتغيرين التنبؤيين قد فسرا حوالى ٩٧,٥٪ من التغير فى ص وهذا جزء كبير يدعم الحكم بوجود العلاقة الخطية .

(حـ) اختبار دلالة كل من معاملى الانحدار :

أعطانا الحاسب فى السطرين السادس والسابع من العمود الثالث التقديرين الخاصين بالخطأ المعيارى لمعاملى الانحدار الجزئيين وهما ع $(ب_1)$ = ٠,١٧٧٠٨ ، ع $(ب_2)$ = ٠,١٦١٢٥ ، كما أعطانا فى العمودين الأخيرين من هذين السطرين قيمتى $ت_1$ ، $ت_2$ مستخدما الصيغة (١٠) بعد وضع $\beta_1 = 0$ ، $\beta_2 = 0$ وهما $ت_1$ = ١١,٨٨٨٩١ ، $ت_2$ = ٥,٤٢٦٥٢ ولكل منهما سبع درجات حرية .

الاستنتاج :

نظرا لأن $ت_{٠,٠١ [٧]}$ = ٣,٤٩٩ نرفض كلا من الفرضين الصفريين $\beta_1 = 0$ ، $\beta_2 = 0$ عند مستوى الدلالة ٠,٠١ وهذا يعنى أن كلا من المتغيرين التنبؤيين يسهم إسهاما جوهريا فى التنبؤ بعدد الوحدات المباعة .

نلخص ما وجدناه فى هذه التجربة كما يلى : بناء على البيانات المشاهدة فى العينة على أساس أن افتراضات الانحدار متحققة ، تأخذ معادلة الانحدار الشكل الآتى :

$$\hat{ص} = ٦,٠٥٣ + ٢,١٠٥\ س_1 + ٠,٨٧٧\ س_2$$

وهذه المعادلة تعبر تعبيرا مناسبا عن العلاقة الحقيقية بين المتغير ص (عدد الوحدات المباعة) والمتغيرين $س_1$ ، $س_2$ (عدد البائعين وقيمة تكاليف الإعلان) ، وتفسر حوالى ٩٧,٥٪ من التغير فى قيم ص . كما أن كلا من هذين المتغيرين يسهم إسهاما جادا فى التنبؤ بقيم هذا المتغير .

(ثانيا) الحل بغير استخدام الحاسب :

إذا لم يكن الحاسب الآلى متوفرا وكان الانحدار ذا متغيرين تنبؤيين فقط ، يمكن إجراء العمليات الحسابية المطلوبة باستخدام حاسب الجيب دون تحمل مشقة كبيرة . وفى هذا المثال يتخذ الحل الخطوات الآتية :

(ا) إيجاد معادلة الانحدار :

نبدأ بحساب المجاميع الآتية

مج ص = ٢٥٠ ، مج $ص^2$ = ٦٣٩٤ ، مج $س_1$ = ٤٠ ، مج $س_1^2$ = ١٨٠ ، مج $س_2$ = ١٢٠ ، مج $س_2^2$ = ١٤٦٤ ، مج $س_1$ $س_2$ = ٤٨٩ ، مج $س_1$ ص = ١٠٥٠ ، مج $س_2$ ص = ٣٠٤٠ ثم نعوض بهذه القيم فى المعادلات المعتادة (٣) كالآتى :

$$١٠\alpha + ٤٠\beta_1 + ١٢٠\beta_2 = ٢٥٠$$

$$٤٠\alpha + ١٨٠\beta_1 + ٤٨٩\beta_2 = ١٠٥٠$$

$$١٢٠\alpha + ٤٨٩\beta_1 + ١٤٦٤\beta_2 = ٣٠٤٠$$

وهناك عدة طرق لحل مثل هذه المعادلات نختار منها هنا طريقة دوليتل Doolittle التى تحول مصفوفة المعاملات والثوابت فى المعادلات المعتادة إلى مصفوفة مثلثية تكون جميع عناصر القطر الرئيسى فيها مساوية للواحد . ولهذه الطريقة روتين خاص من التعليمات تتبين من الجدول (١٢ – ٥) الآتى .

الجدول (١٢ – ٥)

طريقة دولیتل لحل المعادلات الخطية

التعليمات	الصف	α	β_1	β_2	الثوابت
معاملات وثابت المعادلة الأولى	$م_1$	١٠	٤٠	١٢٠	٢٥٠
معاملات وثابت المعادلة الثانية	$م_2$	٤٠	١٨٠	٤٨٩	١٠٥٠
معاملات وثابت المعادلة الثالثة	$م_3$	١٢٠	٤٨٩	١٤٦٤	٣٠٤٠
$م_1$	$م'_1$	١٠	٤٠	١٢٠	٢٥٠
$م'_1 \div ١٠$	$ص_1$	١	٤	١٢	٢٥
$م_{2ر} - م'_{21}\, ص_{1ر}$	$م'_2$	٠	٢٠	٩	٥٠
$م'_2 \div ٢٠$	$ص_2$	٠	١	٠,٤٥	٢,٥
$م_{3ر} - م'_{31}\, ص_{1ر} - م'_{32}\, ص_{2ر}$	$م'_3$	٠	٠	١٩,٩٥	١٧,٥
$م'_3 \div ١٩,٩٥$	$ص_3$	٠	٠	١	٠,٨٧٧١٩

يلاحظ فى هذه التعليمات أن هناك ثلاثة مقادير ثابتة هى $م'_{21} = ٤٠$ ، $م'_{31} = ١٢٠$ ، $م'_{32} = ٩$ وهذة تسمى **محاور الارتكاز** . أما بقية القيم التى تشتمل على الرمز ر فتتغير حسب قيمة ر التى تعبر عن رقم العمود (ر = ١ ، ٢ ، ٣ ، ٤) فمثلا $م_{2ر}$ تعبر عن العنصر الذى بالصف $م_2$ والعمود ر وبهذا يكون

$م_{12} = ٤٠$ ، $م_{22} = ١٨٠$ ، $م_{32} = ٤٨٩$ ، $م_{42} = ١٠٥٠$

تبدأ طريقة دولتيل بتدوين معاملات وثوابت المعادلات كما هو مبين بالصفوف الثلاثة الأولى المشار إليها بالرموز $م_١$ ، $م_٢$ ، $م_٣$. ثم تحسب عناصر الصفوف التالية الواحد بعد الآخر باستخدام التعليمات المبينة أمامه .

فالصف الرابع $م'_١$ يتكون من نفس عناصر الصف $م_١$. والصف التالى $ص_١$ يتكون بقسمة عناصر الصف السابق $م'_١$ على معامل α وهو ١٠ ، والصف التالى $م'_٢$ تتكون عناصره باستخدام التعليمات المبينة أمامه كالآتى :

بوضع ر = ١ نجد أن العنصر الأول فى هذا الصف هو ٤٠ − ٤٠ × ١ = ٠

وبوضع ر = ٢ نجد أن العنصر الثانى فى هذا الصف هو ١٨٠ − ٤٠ × ٤ = ٢٠

وبوضع ر = ٣ نجد أن العنصر الثالث فى هذا الصف هو

٤٨٩ − ٤٠ × ١٢ = ٩

وبوضع ر = ٤ نجد أن العنصر الرابع فى هذا الصف هو

١٠٥٠ − ٤٠ × ٢٥ = ٥٠

ويلاحظ أننا استخدمنا هنا محور الارتكاز $م'_{٢١}$ = ٤٠ .

بنفس الطريقة نوجد بقية الصفوف الثلاثة . وبذلك تتحول المصفوفة التى رمز لصفوفها بالرموز $م_١$ ، $م_٢$ ، $م_٣$ وهى المصفوفة الأصلية إلى المصفوفة المثلثية التى رمز لصفوفها بالرموز $ص_١$ ، $ص_٢$ ، $ص_٣$ وبالتالى تتحول مجموعة المعادلات المعتادة إلى المجموعة المكافئة الآتية :

α + ٤ β_1 + ١٢ β_2 = ٢٥ (من الصف $ص_١$)

، β_1 + ٠,٤٥ β_2 = ٢,٥ (من الصف $ص_٢$)

، β_2 = ٠,٨٧٧١٩ (من الصف $ص_٣$)

من هذه المعادلات ينتج أن القيم أ ، $ب_١$ ، $ب_٢$ التى تحقق هذه المعادلات معا هى :

$ب_٢ = ٠,٨٧٧١٩$

$\therefore ب_١ = ٢,٥ - ٠,٤٥ ب_٢ = ٢,٥ - ٠,٤٥ \times ٠,٨٧٧١٩$

$= ٢,١٠٥٢٦$

$، أ = ٢٥ - ١٢ ب_٢ - ٤ ب_١$

$= ٢٥ - ١٢ \times ٠,٨٧٧١٩ - ٤ \times ٢,١٠٥٢٦$

$= ٦,٠٥٢٦٨$

فتكون معادلة الانحدار مقربة إلى ثلاث خانات عشرية هى :

$$\hat{ص} = ٦,٠٥٣ + ٢,١٠٥ س_١ + ٠,٨٧٧ س_٢$$

وهى نفس المعادلة التى أوجدها لنا الحاسب الآلى .

(ب) اختبار دلالة الانحدار :

نحتاج هنا إلى تحليل الاختلاف فى ص إلى مركبتين كالآتى :

$$\text{الاختلاف الكلى} = م ٢ (ص) = مج ص^٢ - \frac{(مج ص)^٢}{ن}$$

$$= ٦٣٩٤ - \frac{(٢٥٠)^٢}{١٠} = ١٤٤ \quad \text{بدرجات حرية } ن - ١ = ٩$$

من الصيغة (٥) نجد أن :

$$\text{الاختلاف غير المفسر} = مج ص^٢ - أ مج ص - ب_١ مج س_١ص - ب_٢ مج س_٢ص$$

$$= ٦٣٩٤ - ٦,٠٥٢٦٨ \times ٢٥٠ - ٢,١٠٥٢٦ \times ١٠٥٠ - ٠,٨٧٧١٩ \times ٣٠٤٠$$

$$= ٣,٦٤٩٤ \quad \text{بدرجات حرية } ن - ك - ١ = ٧$$

∴ الاختلاف المفسر = ١٤٤ − ٣,٦٤٩٤

= ١٤٠,٣٥٠٦ بدرجات حرية ك = ٢

لاختبار دلالة الانحدار نستخدم الصيغة (٦) كالآتى :

$$ف_ى = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}} = \frac{١٤٠,٣٥٠٦ \div ٢}{٣,٦٤٩٤ \div ٧}$$

$$= ١٣٤,٦٠٤٩$$

وهذه هى نفس القيمة التى أوجدها الحاسب (مع فارق التقريب) وهى كما ذكرنا ذات دلالة عالية وتدعونا لرفض الفرض الصفرى $\beta_1 = \beta_2 = 0$.

$$\text{معامل التحديد } ر^2_{ص(١ ، ٢)} = \frac{\text{الاختلاف المفسر}}{\text{الاختلاف الكلى}} = \frac{١٤٠,٣٥٠٦}{١٤٤}$$

$$= ٠,٩٧٤٦٦$$

وهى نفس القيمة التى أوجدها الحاسب .

(حـ) اختبار دلالة كل من معاملى الانحدار :

نحتاج هنا أولا إلى تقدير الخطأ المعيارى لكل من المعاملين بالصيغة (١١) وهى

$$ع(ب_و) = \frac{ع_{ص.س_١س_٢}}{\sqrt{(١ - ر^2_{٢١}).م م (س_و)}} \qquad و = ١ ، ٢$$

ثم استخدام اختبار ت لكل منهما بالصيغة (١٠) وهى

$$ت_و = \frac{B_و - ٠}{ع(ب_و)} \qquad و = ١ ، ٢$$

بدرجات حرية عددها $\text{ن} - \text{ك} - ١ = ٧$.

لدينا : $\text{ع}^{٢}_{\text{ص.س}_١\text{س}_٢} = \dfrac{١}{\text{ن} - \text{ك} - ١} \times$ الاختلاف غير المفسر

$$= \frac{١}{٧} \times ٣{,}٦٤٩٤ = ٠{,}٥٢١٢٤٢٧$$

$$\therefore \text{ع}_{\text{ص.س}_١\text{س}_٢} = ٠{,}٧٢٢٠$$

$$\text{، م م }(\text{س}_١) = ١٨٠ - \frac{٤٠^{٢}}{١٠} = ٢٠$$

$$\text{، م م }(\text{س}_٢) = ١٤٦٤ - \frac{١٢٠^{٢}}{١٠} = ٢٤$$

$$\text{، م ص }(\text{س}_١ \text{ ، } \text{س}_٢) = ٤٨٩ - \frac{٤٠ \times ١٢٠}{١٠} = ٩$$

$$\therefore \text{ر}^{٢}_{٢١} = \frac{٨١}{٢٠ \times ٢٤} = ٠{,}١٦٨٧٥$$

$$\therefore ١ - \text{ر}^{٢}_{٢١} = ٠{,}٨٣١٢٥$$

ومنها $\text{ع}(\text{ب}_١) = \dfrac{٠{,}٧٢١٩٨}{\sqrt{٠{,}٨٣١٢٥ \times ٢٠}} = ٠{,}١٧٧٠٧$

$$\text{، ع}(\text{ب}_٢) = \frac{٠{,}٧٢١٩٨}{\sqrt{٠{,}٨٣١٢٥ \times ٢٤}} = ٠{,}١٦١٦٤$$

$\therefore \text{ت}_١ = \dfrac{٢{,}١٠٥}{٠{,}١٧٧٠٧} = ١١{,}٨٨٨$ بدرجات حرية ٧

$\text{، ت}_٢ = \dfrac{٠{,}٨٧٧}{٠{,}١٦١٦٤} = ٥{,}٤٢٥٦$ بدرجات حرية ٧

وهاتان القيمتان أعطاهما الحاسب وقد رأينا أن كليهما ذو دلالة عالية وتدعوان إلى رفض كل من الفرضين β_1 = ٠ ، β_2 =٠ ، عند مستوى الدلالة ٠,٠١

(١٢ - ٣) أسلوب آخر لاختبار دلالة معاملات الانحدار الجزئية :

لنبدأ بالحالة التى يكون لدينا فيها متغيران سينيان س~١~ ، س~٢~ . إذا أوجدنا الاختلاف المفسر الناشىء عن انحدار هذين المتغيرين على المتغير ص حين يستخدمان معا (بدرجتين من درجات الحرية) ثم أوجدنا الاختلاف الناشىء عن انحدار المتغير س~٢~ على المتغير ص حين يستخدم وحده (بدرجة واحدة من الحرية) فإن الفرق بين هذين الاختلافين هو مقدار الاختلاف المفسر الذى ساهم به المتغير س~١~ عند ضمه إلى المتغير س~٢~ . أي هو المساهمة الخاصة بالمتغير س~١~ بعد استبعاد مساهمة س~٢~ ويمكن اختبار الفرض الصفرى β_1 = ٠ ضد الفرض β_1 ≠ ٠ بواسطة اختبار ف بالصيغة (٦) كالآتى :

$$ف = \frac{\text{التباين المفسر (للمتغير س}_1\text{ بعد استبعاد س}_2)}{\text{التباين غير المفسر}}$$

$$= \frac{[\text{م م (انحدار س}_1\text{ ، س}_2) - \text{م م (انحدار س}_2)]/1}{[\text{م م(ص)} - \text{م م (انحدار س}_1\text{،س}_2)]/(ن - ٣)} \qquad (١٣)$$

مع ملاحظة أن م م (ص) - م م (انحدار س~١~ ، س~٢~) هو الاختلاف غير المفسر . وبالمثل ، إذا أوجدنا الاختلاف الذى يفسره س~١~ وحده وطرحناه من الاختلاف الذى يفسره المتغيران معا نحصل على المساهمة الخاصة بالمتغير س~٢~ بعد استبعاد أثر س~١~ ويمكن اختبار هذه المساهمة بنفس الطريقة .

ففى المثال – انظر الحل بغير الحاسب – وجدنا ما يلى :

σ^2 (ص) = الاختلاف الكلى = ١٤٤

σ^2 (انحدار $س_1$ ، $س_2$) = الاختلاف المفسر للمتغيرين معا = ١٤٠,٣٥١٢

الاختلاف غير المفسر = ٣,٦٥ بدرجات حرية ن – ٣

نحسب الآن الاختلاف الذى يفسره كل من $س_1$ ، $س_2$ عندما يستخدم كل منهما على حدة .

$$\sigma^2(\text{انحدار } س_1) = \frac{[\sigma^2 \text{ مج } (س_1 ، ص)]^2}{\sigma^2 (س_1)}$$

$$= \frac{\left(١٠٥٠ - \frac{٢٥٠ \times ٤٠}{١٠}\right)^2}{٢٠} = ١٢٥ \text{ بدرجة واحدة من الحرية}$$

$$\sigma^2(\text{انحدار } س_2) = \frac{\left(٣٠٤٠ - \frac{٢٥٠ \times ١٢٠}{١٠}\right)^2}{٢٤} = ٦٦,٦٧ \text{ بدرجة واحدة من الحرية}$$

من الصيغة (١٣) ، لاختبار الفرض β_1 = ٠

$$ف_1 = \frac{١٤٠,٣٥١٢ - ٦٦,٦٧}{٣,٦٥ \div ٧} = \frac{٧٣,٦٨١٢}{٠,٥٢١٤}$$

$= ١٤١,٣١٤$ بدرجتى حرية ١ ، ٧

ولاختبار الفرض β_2 = ٠

$$ف_2 = \frac{١٤٠,٣٥١٢ - ١٢٥}{٣,٦٥ \div ٧} = \frac{١٥,٣٥١٢}{٠,٥٢١٤}$$

$= ٢٩,٤٤٢$ بدرجتى حرية١ ، ٧

ونظرا لأن $ف_{٠,٠١ [١ ، ٧]}$ = ١٢,٢ نرفض كلا من $\beta_{١} = ٠$ ، $\beta_{٢} = ٠$

عند مستوى الدلالة ٠,٠١ .

يلاحظ أن اختبار ف المستخدم هنا يكافىء اختبار ت الذى نتج عن الحل السابق ، ويتأكد هذا من ملاحظة أن $\sqrt{١٤١,٣١٤}$ = ١١,٨٨٨ = $ت_{١}$ ، $\sqrt{٢٩,٤٤٢}$ = ٥,٤٢٦ = $ت_{٢}$. ويمكن إذا أردنا أن ندون هذه النتائج فى جدول كالآتى .

الجدول (١٢ – ٦)

اختبار كل من المتغيرين بعد استبعاد أثر المتغير الآخر

مصدر الاختلاف	م ٢	د ح	ط ٢	ف
انحدار $س_{١}$ ، $س_{٢}$ معا	١٤٠,٣٥	٢		
انحدار $س_{١}$ وحده	١٢٥	١		
انحدار $س_{٢}$ وحده	٦٦,٦٧	١		
انحدار $س_{١}$ بعد $س_{٢}$	٧٣,٦٨	١	٧٣,٦٨	١٤١,٣١٤**
انحدار $س_{٢}$ بعد $س_{١}$	١٥,٣٥	١	١٥,٣٥	٢٩,٢٤٢**
الاختلاف غير المفسر	٣,٦٥	٧	٠,٥٢١٤	
الكلى	١٤٤,٠٠	٩		

ملاحظات :

من التحليل الملخص بالجدول (١٢ – ٦) نخرج بالنتائج والملاحظات الآتية :

(١) المتغير $س_١$ حين يستخدم وحده للتنبؤ بقيم ص أفضل من المتغير $س_٢$ حين يستخدم وحده لنفس الغرض ، وذلك لأن الاختلاف الذى يفسره $س_١$ منفردا وهو ١٢٥ أكبر من الاختلاف الذى يفسره $س_٢$ منفردا وهو ٦٦,٦٧ . ونصل إلى نفس النتيجة إذا قارنا معاملى التحديد ، فمعامل التحديد للمتغير $س_١$ هو :

$$ر^٢_{ص س_١} = \frac{١٢٥}{١٤٤} = ٠,٨٦٨٠٦ \quad \text{بينما} \quad ر^٢_{ص س_٢} = \frac{٦٦,٦٧}{١٤٤} = ٠,٤٦٢٩٩$$

وبهذا الأسلوب نستطيع مقارنة أى عدد من المتغيرات تستخدم فرادا .

(٢) فى تفسير الاختلاف الكلى فى ص تكون مساهمة أى متغير أكبر فى حالة استخدامه منفردا عنها فى حالة استخدامه بعد متغير آخر فبالنسبة للمتغير $س_١$ نجد أن ١٢٥ > ٧٣,٦٨ وبالنسبة للمتغير $س_٢$ نجد أن ٦٦,٦٧ > ١٥,٣٥ وهذه نتيجة عامة مهما كان عدد المتغيرات التنبؤية ، فالمساهمة المنفردة لأى متغير تكون أكبر دائما من مساهمته مع متغير أو عدة متغيرات أخرى .

(٣) معامل التحديد المحسوب من معادلة الانحدار للمتغيرين $س_١$ ، $س_٢$ معا وهو $ر^٢_{ص (١ ، ٢)} = \frac{١٤٠,٣٥}{١٤٤} = ٠,٩٧٥$ أكبر من معامل التحديد المحسوب من معادلة الانحدار لأى من المتغيرين وهما ٠,٨٦٨ ، ٠,٤٦٣ . وهذه نتيجة عامة ، فمعامل التحديد يقل دائما حين يستبعد واحد أو أكثر من المتغيرات التنبؤية .

(٤) فى معادلة الانحدار للمتغيرين $س_١$ ، $س_٢$ معا يصعب تقدير الأهمية النسبية لهذين المتغيرين من حيث مقدار المساهمة فى التنبؤ بقيم المتغير التابع ص ،

لأننا إذا قدرنا هذه الأهمية النسبية بواسطة الاختلافين المفسرين عند استخدام كل متغير على حدة وهما ١٢٥ ، ٦٦,٦٧ نجد أن هذا التقدير غير مناسب لأن مجموع هذين الاختلافين وهو ١٩١,٦٧ يزيد عن الاختلاف الكلى فى ص وهو ١٤٤ . ومن ناحية أخرى ، إذا اعتبرنا الاختلاف الناشىء عن المتغير $س_{١}$ بعد استبعاد أثر المتغير $س_{٢}$ وهو ٧٣,٦٨ والاختلاف الناشىء عن المتغير $س_{٢}$ بعد استبعاد أثر المتغير $س_{١}$ وهو ١٥,٣٥ نجد أن مجموعهما وهو ٨٩,٠٣ يقل عن الاختلاف الذى يفسره المتغيران معا وهو ١٤٠,٣٥ . وهذه الصعوبة فى تقدير الأهمية النسبية للمتغيرات هى إحدى الصعوبات التى يلقاها الباحث عند التصدى لمقارنة المتغيرات فى الانحدار الخطى المتعدد وعند اختيار أفضل المتغيرات التى تدخل فى معادلة الانحدار كما سيأتى بالبند التالى .

(٥) قمية أى معامل انحدار جزئى $ب_{ى}$ لمتغير $س_{ى}$ هى قيمة شرطية تختلف باختلاف المتغيرات التى تدخل معه فى معادلة الانحدار .

إن أسلوب اختبار دلالة معاملات الانحدار الجزئية الذى أدى إلى الصيغة (١٣) هو أسلوب عام مهما كان عدد المتغيرات التنبؤية ، ويستخدم فى التعرف على دلالة ما يحدث من نقص فى دقة التنبؤ حين يستبعد واحد أو أكثر من المتغيرات من معادلة الانحدار . وبصفة عامة إذا كنا قد وفقنا معادلة انحدار فى ك من المتغيرات $س_{١}$ ، $س_{٢}$ ، ... ، $س_{ك}$ ووفقنا معادلة انحدار فى ل < ك من هذه المتغيرات أى بعد استبعاد ك − ل منها ، وأردنا اختبار ما إذا كانت دقة التنبؤ قد نقصت نقصا ذا دلالة نتيجة لهذا الاستبعاد فإننا نستخدم الصيغة العامة الآتية :

$$ف = \frac{[م^{٢}(ك) - م^{٢}(ل)]/(ك - ل)}{[م^{٢}(ص) - م^{٢}(ك)]/(ن - ك - ١)} \qquad (١٤)$$

بدرجتى حرية ك − ل ، ن − ك − ١ ،

حيث م ٢ (ك) هو الاختلاف الناشىء عن الانحدار على ك من المتغيرات
، م ٢ (ل) هو الاختلاف الناشىء عن الانحدار على ل من هذه المتغيرات
، م ٢ (ص) هو الاختلاف فى قيم المتغير التابع ص
إذ يمكن إثبات أن كلا من بسط ومقام هذه النسبة هو تقدير مستقل للتباين $\sigma^{٢}$.

إن الصيغة (١٤) يمكن كتابتها بدلالة معاملات التحديد كالآتى :

$$\text{ف} = \frac{[\text{ر}^{٢}(\text{ك}) - \text{ر}^{٢}(\text{ل})] \;/\; (\text{ك} - \text{ل})}{[١ - \text{ر}^{٢}(\text{ك})] \;/\; (\text{ن} - \text{ك} - ١)} \qquad (١٥)$$

بدرجتى حرية (ك − ل) ، (ن − ك − ١)

مثال (١٢ − ٢) :

أجرى انحدار خطى لمتغير عشوائى ص على خمسة متغيرات تنبؤية $\text{س}_{١}$ ، $\text{س}_{٢}$ ، $\text{س}_{٣}$ ، $\text{س}_{٤}$ ، $\text{س}_{٥}$ ووجد أن معامل التحديد فى عينة حجمها ٤٦ هو ٠,٦٢ وعندما أجرى انحدار خطى لنفس المتغير على ثلاثة من هذه المتغيرات $\text{س}_{١}$ ، $\text{س}_{٣}$ ، $\text{س}_{٤}$ وجد أن معامل التحديد ٠,٥٧ . هل يمكن الاستغناء عن المتغيرين $\text{س}_{٢}$ ، $\text{س}_{٥}$ والاكتفاء بالمتغيرات الثلاثة الأخرى دون أن نفقد شيئا من دقة التنبؤ بقيم المتغير ص ؟

الحل :

من الصيغة (١٥) :

$$\text{ف}_{\text{ى}} = \frac{(٠,٦٢ - ٠,٥٧) \div ٢}{(١ - ٠,٦٢) \div ٤٠} = ٢,٦٣٢$$

وهذه القيمة تقل عن القيمة الحرجة $ف_{٠,٠٥[٢،٤٠]} = ٣,٢٣$ مما يدعونا إلى قبول الفرض الصفرى أن دقة التنبؤ لم تتأثر ، وبالتالى نستطيع الاكتفاء بالمتغيرات $س_1$ ، $س_3$ ، $س_4$ للتنبؤ بالمتغير ص دون أن نفقد شيئا من دقة التنبؤ .

مثال (١٢ – ٣) :

فى المثال (١٢ – ١) هل نستطيع الاكتفاء بأحد المتغيرين للتنبؤ بقيم المتغير التابع ص دون أن نفقد شيئا من دقة التنبؤ ؟

الحل :

سبق أن أجبنا عن هذا السؤال حين استخدمنا الصيغة (١٣) – التى هى حالة خاصة من الصيغة (١٤) أو (١٥) — فى رفض كل من الفرضين $\beta_1 = ٠$ ، $\beta_2 =$ لأن هذا يعنى أن وجود أى من المتغيرين له أثر ذو دلالة فى عملية التنبؤ . باستخدام الصيغة (١٥) لاختبار إمكانية الاستغناء عن المتغير $س_1$ لدينا :

$$ر^2_{ص(س_1،س_2)} = ٠,٩٧٤٦٦ \quad ، \quad ر^2_{ص س_2} = ٠,٤٦٢٩٩$$

$$ف_{ى} = \frac{(٠,٩٧٤٦٦ - ٠,٤٦٢٩٩) \div ١}{(١ - ٠,٩٧٤٦٦) \div ٧} = \frac{٠,٥١١٦٧}{٠,٠٠٣٦٢}$$

$$= ١٤١,٣٥$$

وهذه القيمة ذات دلالة عالية وتعنى أن الاستغناء عن $س_1$ يقلل من دقة التنبؤ . وبالمثل ، لاختبار إمكانية الاستغناء عن $س_2$

$ف = \frac{٠,٩٧٤٦٦ - ٠,٨٦٨٠٦}{٠,٠٠٣٦٢} = ٢٩,٤٥$ وهذه أيضا ذات دلالة عالية وتعنى أن الاستغناء عن $س_2$ يقلل من دقة التنبؤ .

(١٢ - ٤) اختيار متغيرات التنبؤ :

من الصعوبات التى يلقاها الباحثون فى الانحدار المتعدد كيفية اختيار المتغيرات التى تدخل فى معادلة الانحدار لتعطى أعلى درجة من الدقة فى التنبؤ بالمتغير التابع ص . وفى المعتاد يكون لدى الباحث مجموعة كبيرة من المتغيرات المرشحة لذلك ويكون من المفيد عمليا اختصار هذه المجموعة إلى مجموعة تتكون من أقل عدد ممكن من هذه المتغيرات بشرط أن تعطى هذه المجموعة الجزئية معادلة تنبؤ لا تقل كفاءة عن المعادلة التى تستخدم جميع المتغيرات المتاحة . إن عملية الاختصار هذه ممكنة فى الغالب لأن بعض المتغيرات المتاحة لا تسهم بدرجة كافية فى عملية التنبؤ كما قد يبدو لأول وهلة ، كما أنه قد توجد مجموعة (أو أكثر) من المتغيرات المتاحة ترتبط ببعضها ارتباطا عاليا وبالتالى تعطى معلومات مماثلة ويمكن حينئذ الاكتفاء بمتغير واحد فقط من هذه المجموعة يمثلها جميعا .

ويتلخص السؤال المطروح هنا فيما يلى : إذا كان لدينا ك من المتغيرات التنبؤية المرشحة للدخول فى معادلة الانحدار فكيف نختار أصغر مجموعة جزئية من هذه المجموعة بحيث تعطى نفس الدرجة من الدقة فى التنبؤ بالمتغير التابع ؟ فمثلا إذا كان لدينا مجموعة من عشرة متغيرات تنبئوية فهل يمكن أن نكتفى بمجموعة من اثنين أو ثلاثة منها لتعبر عن المجموعة بكاملها (دون أن نفقد شيئا من دقة التنبؤ) ؟

الواقع أنه لا توجد إجابة واحدة شافية لهذا السؤال ولكن هناك عدة اقتراحات بمحاولات نسترشد بها فى تحديد المجموعة المطلوبة ، وتتطلب هذه المحاولات النظر إلى البيانات المشاهدة عدة مرات من جوانب مختلفة ، وهنا يلعب الحاسب دورا رئيسيا لأن تنفيذ هذه المحاولات يحتاج إلى جهد حسابى شاق .

نفرض أننا حصلنا على بيانات مشاهدة فى عينة حجمها ن كما فى الجدول (١٢ - ١) السابق .

(ا) إن أول ما يخطر بالبال هو اختيار المتغيرات الأكثر علاقة بالمتغير التابع ص ، فنقوم بالمقارنة بينها كمتغيرات فرادا إما عن طريق مقادير الاختلافات المفسرة م م (الانحدار) أو عن طريق معاملات التحديد $ر^{٢}$ كما جاء بالملاحظة (١) بالبند (١٢ - ٣) السابق ، فتكون المتغيرات ذوات القيم الأكبر مرشحة مبدئيا للدخول فى معادلة انحدار متعدد .

(ب) نبحث عما إذا كان هناك مجموعة (أو أكثر) من المتغيرات التنبؤية التى ترتبط ببعضها ارتباطا عاليا ، فإذا وجدت مثل هذه المجموعة فإن أحدها يمكن أن يمثلها جميعا . ويحتاج الأمر هنا إلى إيجاد معامل الارتباط بين كل زوج من هذه المتغيرات ، ويفضل وضع هذه المعاملات وأيضا معاملات الارتباط بين كل من المتغيرات والمتغير التابع فى جدول لتسهيل الرجوع إليها . وعلى فرض أن هناك خمس متغيرات تنبؤية يكون الجدول كالآتى :

	$س_{١}$	$س_{٢}$	$س_{٣}$	$س_{٤}$	$س_{٥}$
$س_{١}$	١	$ر_{٢١}$	$ر_{٣١}$	$ر_{٤١}$	$ر_{٥١}$
$س_{٢}$		١	$ر_{٣٢}$	$ر_{٤٢}$	$ر_{٥٢}$
$س_{٣}$			١	$ر_{٤٣}$	$ر_{٥٣}$
$س_{٤}$				١	$ر_{٥٤}$
ص	$ر_{ص١}$	$ر_{ص٢}$	$ر_{ص٣}$	$ر_{ص٤}$	$ر_{ص٥}$

(حـ) نوجد معادلة انحدار ص على جميع ما لدينا من متغيرات ثم نفحص دلالة كل من معاملات الانحدار الجزئية $ب_{١}$ ، $ب_{٢}$ ، . . . ، $ب_{ك}$ عن طريق اختبار

ت بالصيغة (١٠) فتكون المتغيرات التى ترشح للدخول فى معادلة الانحدار هى تلك التي تحظى بالقيم الأكبر من قيم ت ، أما المتغيرات التى تناظر القيم الأصغر من ت فتكون معرضة للاستبعاد . على أن ذلك لا ينبغى أن يتم دون تدبر فإن استخدام أو استبعاد متغير مالا يقرر لمجرد أن قيمة ت المناظرة كبيرة أو صغيرة ، لأن قيم معاملات الانحدار الجزئية وبالتالى قيم ت هى قيم شرطية وتتغير بتغير المتغيرات التى تدخل معها فى الانحدار كما سبق القول بالملاحظة (٥) ، ومن ناحية أخرى قد توجد أسباب نظرية أو خبرات تحتم الاحتفاظ ببعض المتغيرات حتى ولو كان معامل الانحدار الجزئى المناظر غير ذى دلالة . فمثلا إذا كان لدينا عدة متغيرات نرغب استخدامها فى التنبؤ بالمسافة التى تقطعها سيارة ما وكان من بين هذه المتغيرات متغير « مقدار الوقود المستهلك » فلابد من الاحتفاظ بهذا المتغير حتى ولو كان معامل الانحدار المناظر له غير ذى دلالة .

(ى) على ضوء ما نجده فى (أ) و(ب) و(حـ) نختار مجموعة أو أكثر من المتغيرات ونحسب معادلات انحدار كل منها مع إيجاد معامل التحديد ومقارنته بمعامل التحديد الذى ينجم من الانحدار على جميع المتغيرات المتاحة واختبار دلالة الفرق بينهما فى دقة التنبؤ باستخدام الصيغة (١٥) . وإذا وجدنا أن إحدى المجموعات المختارة بنفس كفاءة المجموعة الكاملة نحاول اختزالها بنفس الطريقة إلى مجموعة ذات عدد أقل من المتغيرات .

ويبنى اختيارنا النهائى لمجموعة المتغيرات التى تدخل فى معادلة التنبؤ التى ننشدها على أساس أنها (أولا) تشتمل على أقل عدد من المتغيرات و(ثانيا) يمكنها أن تعبر عن مجموعة المتغيرات المتاحة بكاملها ، أى بحيث لا تقل دقة التنبؤ منها عن دقة التنبؤ من استخدام جميع المتغيرات المتاحة . هذا ويجدر الإشارة أن المجموعة التى تختار على هذا الأساس ليست فريدة ، بل يمكن أن نعثر على أكثر من مجموعة تستوفى الشرطين المطلوبين .

تمارين (١٢ - ١)

أجريت دراسة لمعرفة أثر العوامل الجغرافية على حجم نوع من الضفادع واستهدف جزء من هذه الدراسة التنبؤ بطول جسم الضفدعة عن طريق الارتفاع عن سطح البحر والمتوسط السنوى لدرجة الحرارة : وقد جمعت بيانات من ١٣ موقعا ودونت بالجدول الآتى :

الجدول (١٢ - ٧)

الموقع	طول الضفدعة ص	الارتفاع عن سطح البحر س١	درجة الحرارة س٢
(١)	٢٣,٧	٦٠٠	٦٠
(٢)	٢٣,٩	٨٠٠	٥٥
(٣)	٢٣,٤	١٤٥٠	٥٠
(٤)	٢٣,٤	١٤٠٠	٥٠
(٥)	٢٢,٢	١٠٠	٥٥
(٦)	١٩,٥	٤٥٠	٥٠
(٧)	٢٣,٥	٤٠٠	٥٠
(٨)	٢٤,٧	٨٠٠	٦٠
(٩)	٢٣,٣	٨٠٠	٦٠
(١٠)	٢١,٧	١١٠٠	٥٠
(١١)	٢١,٣	٥٠٠	٥٠
(١٢)	٢٠,٢	٤٠٠	٥٠
(١٣)	٢٢,١	٤٠٠	٥٠

مجـ ص = ٢٩٢,٩   مجـ ص² = ٦٦٢٧,٣٧   مجـ س١ = ٩٢٠٠   مجـ س١² = ٨٤٩٥٠٠٠

مجـ س٢ = ٦٩٠   مجـ س٢² = ٣٦٨٥٠   مجـ س١ س٢ = ٤٨٦٥٠٠

مجـ ص س١ = ٢١٠٢٦٥   مجـ ص س٢ = ١٥٥٩٢,٥

(أولا) استخدم مخرجات الحاسب الالكترونى المبينة بالجدول (١٢ – ٨) الآتى لإيجاد معادلة الانحدار الخطي للمتغير ص على المتغيريـن س١ ، س٢ .

(ثانيا) اختبر دلالة الانحدار ككل ودلالة كل من معاملى الانحدار . هل من الممكن الاستغناء عن أحد المتغيرين التنبؤيين دون أن نفقد شيئا من دقة التنبؤ ؟

(ثالثا) أجب عن الجزئين (أولا) و(ثانيا) دون الاستعانة بمخرجات الحاسب .

الجدول (١٢ – ٨)

Regress. of	Y	Length of Frog		
On	X_1	Altitude		
	X_2	Temperature		
Variable Name	Regress. Coeff.	S.E. of Coeff.	t	D.F.
Constant	9.785579	4.129044	2.37	
X_1	.1700912 E-02.	.81158094—03	2.10	10
X_2	.2174479	.7589070 E—1	2.87	10

Multiple Corr. Coeff. (R) = .73379

Coeff. of Determination (R ^ 2) = .5384

Estimated Standard Error of Estimate = 1.1390588

Analyis of Variance for Regressioon

Source of Variation	SS	D.F.	M S	F
Regression	15.13315	2	7.566575	5.8319
Residual	12.97455	10	1.297455	

ملاحظة :

الرمز E - 01 يعنى ضرب العدد الذى على يسار هذا الرمز فى $١٠^{-١}$

الرمز E - 02 يعنى ضرب العدد الذى على يسار هذا الرمز فى $١٠^{-٢}$
وهكذا ...

الرمز E + 02 يعنى ضرب العدد الذى على يسار هذا الرمز فى $١٠^{٢}$
وهكذا ...

فمثلا العدد 1700912 E - 02 . يعنى العدد 001700912 .

هذا ما يسمى بالرمزية العلمية scientific notation

ثانيا – الارتباط الخطى المتعدد

سنفترض هنا أن لدينا ك + ١ من المتغيرات العشوائية $ص_١$ ، $ص_٢$ ، ... ، $ص_{ك+١}$ كما سنفترض أن توزيع الاحتمال المشترك لهذه المتغيرات هو توزيع معتدل متعدد المتغيرات multivariate normal distribution إن هذا الفرض يتضمن أمرين : الأول أن توزيع أى من هذه المتغيرات (على حدة) هو توزيع معتدل ، والثانى أن العلاقة بين أى من هذه المتغيرات وأى متغير آخر أو مجموعة من المتغيرات الأخرى هى علاقة خطية .

وامتدادا لدراساتنا فى الارتباط الخطى البسيط فى الفصل العاشر نتناول فى البندين الآتيين معاملين رئيسيين هما معامل الارتباط الخطى المتعدد ومعامل الارتباط الجزئى ونرى كيف نختبر دلالة كل منهما .

(١٢ – ٥) معامل الارتباط المتعدد

MULTIPLE CORRELATION COEFFICIENT

إذا أردنا تقدير درجة العلاقة الخطية بين أحد هذه المتغيرات وليكن $ص_١$ وبقية المتغيرات ، فإننا نعرّف معامل الارتباط المتعدد بين $ص_١$ والمتغيرات الأخرى بأنه

معامل الارتباط البسيط بين المتغير $ص_١$ والمتغير العشوائى الذى يتكون من التركيب الخطى $\alpha + \beta_٢ ص_٢ + \beta_٣ ص_٣ + \dots + \beta_{ك+١} ص_{ك+١}$ وسنرمز لهذا المعامل بالرمز $ر_{١(٢٣ \dots ك+١)}$.

ولتقدير هذا المعامل نأخذ عينة عشوائية من ن من وحدات المجتمع ونقوم بقياس قيمة كل من هذه المتغيرات لكل وحدة فنحصل على ن من المشاهدات كل منها على الصورة ($ص_{ر١}$ ، $ص_{ر٢}$ ، ... ، $ص_{ر، ك+١}$) حيث ر = ١ ، ٢ ، ... ، ن وتوضع هذه القياسات عادة كما فى الجدول الآتى .

جدول (١٢ – ٨)

المشاهدات	$ص_١$	$ص_٢$	$ص_٣$	...	$ص_{ك+١}$
(١)	$ص_{١١}$	$ص_{٢١}$	$ص_{٣١}$	...	$ص_{١، ك+١}$
(٢)	$ص_{١٢}$	$ص_{٢٢}$	$ص_{٣٢}$	...	$ص_{٢، ك+١}$
...	...	...	...	...	...
(ر)	$ص_{ر١}$	$ص_{ر٢}$	$ص_{ر٣}$	...	$ص_{ر، ك+١}$
...	...	...	...	...	...
(ن)	$ص_{ن١}$	$ص_{ن٢}$	$ص_{ن٣}$	...	$ص_{ن، ك+١}$

من هذه المشاهدات نحصل على التقدير المطلوب بعدد سنرمز له بالرمز $ر_{١(٢٣ \dots ك+١)}$ ونحسبه كما فى حالة الارتباط الخطى البسيط من الصيغة .

$$ر^٢_{١(٢٣ \dots ك)} = \frac{\text{الاختلاف المفسر}}{\text{الاختلاف الكلى}} \qquad (١٦)$$

أى أن التقدير المطلوب هو الجذر التربيعى لنسبة الاختلاف الذى يفسره الانحدار الخطى المتعدد إلى الاختلاف الكلى فى المتغير $\text{ص}_{\text{١}}$. ويمكن إثبات أنه فى هذه الحالة .

$$\text{١} \geq \text{ص}_{\text{١ (٢٣ ... ك + ١)}} \geq \text{٠} \qquad (\text{١٧})$$

أى أن هذا المعامل لا يمكن أن يكون سالبا وهو يختلف فى هذه الصفة عن معامل الارتباط البسيط الذى نعلم أنه يمكن أن يأخذ قيما سالبة .

وكما فى الانحدار الخطى المتعدد نعتمد فى حساب هذا المعامل على الحاسب الالكترونى توفيرا للجهد والوقت . وفى الحالة البسيطة التى يكون لدينا فيها ثلاثة متغيرات عشوائية $\text{ص}_{\text{١}}$ ، $\text{ص}_{\text{٢}}$ ، $\text{ص}_{\text{٣}}$ يمكننا حساب معامل الارتباط المتعدد $\text{ر}_{\text{١ (٢ ، ٣)}}$ بين المتغير $\text{ص}_{\text{١}}$ والمتغيرين $\text{ص}_{\text{٢}}$ ، $\text{ص}_{\text{٣}}$ من الصيغة الآتية :

$$\text{ر}^{\text{٢}}_{\text{١ (٣٢)}} = \frac{\text{ر}^{\text{٢}}_{\text{٢١}} + \text{ر}^{\text{٢}}_{\text{٣١}} - \text{٢}\ \text{ر}_{\text{٢١}} \cdot \text{ر}_{\text{٣١}} \cdot \text{ر}_{\text{٣٢}}}{\text{١} - \text{ر}^{\text{٢}}_{\text{٣٢}}} \qquad (\text{١٨})$$

ويتطلب حساب هذه القيمة إيجاد كل من :

$\text{ر}_{\text{٢١}}$ وهو معامل الارتباط البسيط بين المتغيرين $\text{ص}_{\text{١}}$ ، $\text{ص}_{\text{٢}}$

، $\text{ر}_{\text{٣١}}$ وهو معامل الارتباط البسيط بين المتغيرين $\text{ص}_{\text{١}}$ ، $\text{ص}_{\text{٣}}$

، $\text{ر}_{\text{٣٢}}$ وهو معامل الارتباط البسيط بين المتغيرين $\text{ص}_{\text{٢}}$ ، $\text{ص}_{\text{٣}}$

ومن الواضح أن قيمة معامل الارتباط المتعدد تعتمد على قيم معاملات الارتباط البسيط بين أزواج المتغيرات المستخدمة ، وهذه الملاحظة صحيحة مهما كان عدد المتغيرات .

اختبار دلالة معامل الارتباط المتعدد :

لاختبار الفرض الصفرى $ر_{١(٣٢ ... ك + ١)} = ٠$ ضد الفرض $ر_{١(٣٢ ... ك +١)} \neq ٠$
نستخدم نفس الصيغة (٦) التى وردت بالبند (١٢ – ١) وهى :

$$ف = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}}$$

$$= \frac{ر^٢_{١(٣٢ ... ك + ١)} / ك}{(١ - ر^٢_{١(٣٢ ... ك + ١)}) / (ن - ك - ١)} \qquad (١٩)$$

وإذا توفرت الشروط المذكور فى مستهل الجزء الثانى من هذا الفصل فإن هذه الإحصاءة يكون لها توزيع ف بدرجتى حرية ك ، ن – ك – ١ حيث ك + ١ عدد جميع المتغيرات المستخدمة . يلاحظ أن كلا من الصيغتين (٦) ، (١٩) تختبر وجود أو عدم وجود العلاقة الخطية .

مثال (١٢ – ٤) :

فى عينة عشوائية من ٢٣ مجموعة من القيم من مجتمع معتدل ذى ثلاثة متغيرات عشوائية وجدت معاملات الارتباط البسيطة الآتية : $ر_{٢١} = ٠,٣٨$ ، $ر_{٢١} = ٠,٤٦$ ، $ر_{٣٢} = ٠,٥١$. أوجد معامل الارتباط المتعدد $ر_{١(٣٢)}$ واختبر دلالته .

الحل :

من الصيغة (١٨) :

$$ر^٢_{١(٣٢)} = \frac{٠,١٤٤٤ + ٠,٢١١٦ - ٢ \times ٠,٣٨ \times ٠,٤٦ \times ٠,٥١}{١ - ٠,٢٦٠١}$$

$$= ٠,٢٤٠$$

∴ معامل الارتباط المتعدد المطلوب = $\sqrt{٠,٢٤٠}$ = ٠,٤٩ تقريبا .

من الصيغة (١٩) :

$$ف_{ى} = \frac{٠,٢٤٠ \div ٢}{(١ - ٠,٢٤) \div ٢٠} = ٣,١٥٧٩ \quad \text{بدرجتي حرية ٢،٢٠}$$

وهذه القيمة ليست بذات دلالة عند المستوى ٠,٠٥ وتشير إلى عدم وجود ارتباط خطي بين المتغير $ص_{١}$ والمتغيرين $ص_{٢}$ ، $ص_{٣}$.

(١٢ - ٦) معامل الارتباط الجزئي

PARTIAL CORRELATION COEFFICIENT

إن معامل الارتباط البسيط $ص_{١٢}$ بين متغيرين عشوائيين $ص_{١}$ ، $ص_{٢}$ يفترض أن هذين المتغيرين يتأثران ببعضهما فقط ولا يتأثران بأي متغير آخر . إلا أنه في كثير من الحالات يكون كل من هذين المتغيرين متأثرا تأثرا خطيا بمتغير عشوائي ثالث $ص_{٣}$ ، وفي هذه الحالة يقتضي القياس الصحيح للارتباط بين $ص_{١}$ ، $ص_{٢}$ استبعاد أثر هذا المتغير الثالث من كل منهما . إن المعامل الذي يقيس مثل هذا الارتباط يسمى بمعامل الارتباط الجزئي بين $ص_{١}$ ، $ص_{٢}$ وسنرمز له بالرمز $ص_{١٢-٣}$. ويعرّف هذا المعامل أيضا بأنه معامل الارتباط البسيط بين $ص_{١}$ ، $ص_{٢}$ عند قيمة ثابتة للمتغير $ص_{٣}$ أي في قطاع الوحدات التي تأخذ نفس القيمة للمتغير $ص_{٣}$. ويتفق هذان التعريفان طالما كان توزيع الاحتمال المشترك للمتغيرات الثلاثة معتدلا trivariate normal distribution إذ أنه في هذه الحالة تظل قيمة $ص_{١٢-٣}$ كما هي مهما كانت القيمة الثابتة للمتغير $ص_{٣}$ ، وهذه هي الحالة التي نتناولها هنا والتي تتمشى مع ما افترضناه في مستهل هذا الجزء .

ويقدر هذا المعامل من العينة بمقدار سنرمز له بالرمز $ر_{١٢-٣}$ حيث

$$ر_{١٢-٣} = \frac{ر_{١٢} - ر_{١٣} \cdot ر_{٢٣}}{\sqrt{(١ - ر^{٢}_{١٣})(١ - ر^{٢}_{٢٣})}} \quad (٢٠)$$

وحيث $ر_{١٢}$ ، $ر_{١٣}$ ، $ر_{٢٣}$ هي معاملات الارتباط البسيط بين أزواج المتغيرات . وبنفس الطريقة نحسب كلا من $ر_{١٣-٢}$ ، $ر_{٢٣-١}$ مع تذكر أننا هنا لا نميز بين متغير مستقل ومتغير تابع .

أما دلالة هذا المعامل فنختبره عن طريق الإحصاءة :

$$ت = \frac{ر_{١٢-٣}\sqrt{ن - ٣}}{\sqrt{١ - ر^{٢}_{١٢-٣}}} \quad \text{بدرجات حرية } ن - ٣ \qquad (٢١)$$

أو عن طريق الإحصاءة المكافئة:

$$ف = \frac{ر^{٢}_{١٢-٣}}{(١ - ر^{٢}_{١٢-٣}) / (ن - ٣)} \quad \text{بدرجات حرية } ١ \text{ ، } ن - ٣ \qquad (٢٢)$$

مثال (١٢ – ٥) :

فى دراسة للعلاقة بين ضغط الدم وكلوسترول الدم أخذت عينة عشوائية من ١٤٢ من السيدات كبار السن وقيس كل من ضغط الدم $ص_{١}$ ودرجة تركيز الكلوسترول $ص_{٢}$ والعمر $ص_{٣}$ ، ثم حسبت معاملات الارتباط البسيط لكل زوج من هذه المتغيرات فوجدت كما يلى : $ر_{١٢} = ٠,٢٤٩٥$ ، $ر_{١٣} = ٠,٣٣٣٢$ ، $ر_{٢٣} = ٠,٥٠٢٩$ ابحث دلالة معامل الارتباط بين ضغط الدم ودرجة تركيز الكلوسترول (أولا) دون استبعاد متغير العمر ، (ثانيا) بعد استبعاد متغير العمر .

الحل :

(أولا) معامل الارتباط بين ضغط الدم ودرجة تركيز الكلوسترول دون

استبعاد متغير العمر هو $ر_{٢١}$ = ٠,٢٤٩٥ ونختبر دلالة هذا المعامل باستخدام الصيغة (٨) بالبند (١٠ — ٥) وهى :

$$ت = \frac{ر_{٢١} \sqrt{ن - ٢}}{\sqrt{١ - ر^{٢}_{٢١}}}$$ بدرجات حرية ن — ٢

أو الصيغة المكافئة ف = $\frac{ر^{٢}_{٢١}}{(١ - ر^{٢}_{٢١}) / (ن - ٢)}$ بدرجتى حرية ١ ، ن — ٢

لدينا : $ت_{ى} = \frac{٠,٢٤٩٥ \sqrt{١٤٠}}{\sqrt{١ - (٠,٢٤٩٥)^{٢}}} = ٣,٠٤٨٦$ بدرجات حرية ١٤٠

وهذه القيمة تزيد عن القيمة الحرجة $ت_{٠,٠١[١٤٠]}$ = ٢,٥٧٦ فهى ذات دلالة عالية وتشير إلى وجود ارتباط خطى بين المتغيرين .

(ثانيا) معامل الارتباط بين ضغط الدم ودرجة تركيز الكلوسترول بعد استبعاد متغير العمر هو معامل الارتباط الجزئى $ر_{٢١-٣}$. من الصيغة (٢٠) نجد أن

$$ر_{٢١-٣} = \frac{٠,٢٤٩٥ - ٠,٣٣٣٢ \times ٠,٥٠٢٩}{\sqrt{(١-٠,٣٣٣٢^{٢})(١-٠,٥٠٢٩^{٢})}} = ٠,١٢٣٣$$

ونختبر دلالة هذا المعامل من الصيغة (٢١) كالآتى :

$$ت_{ى} = \frac{٠,١٢٣٣ \sqrt{١٣٩}}{\sqrt{١ - ٠,١٢٣٣^{٢}}} = ١,٤٦٥$$ بدرجات حرية ١٣٩

وهذه القيمة تقل عن القيمة الحرجة $\text{ت}_{٠,٠٥\,(١٣٩)} = ١,١٩٨٠$ فهى ليست بذات دلالة عند المستوى ٠,٠٥ ، وتشير إلى أنه حين نستبعد عامل العمر لا نستطيع الاستدلال على وجود ارتباط خطى بين ضغط الدم وكلوسترول الدم . هـذا مع ملاحظة أن كلا من ضغط الدم ومقدار الكلوسترول فى الدم يزداد بزيادة العمر ، ومن هنا نرى أن النتيجة التى حصلنا عليها فى (أولا) كانت نتيجة مضللة .

معاملات الارتباط الجزئى من مراتب أعلى :

إن معامل الارتباط الجزئى من النوع $\text{ر}_{\text{س ص}-\text{ل}}$ يوصف بأنه معامل ارتباط جزئى من المرتبة الأولى of the first order لأنه يعطى معامل الارتباط بين متغيرين بعد استبعاد أثر متغير واحد من كل منهما . وبالمثل إذا كان لدينا أربعة متغيرات عشوائية $\text{ص}_١$ ، $\text{ص}_٢$ ، $\text{ص}_٣$ ، $\text{ص}_٤$ ، فإننا نرمز بالرمز $\text{ر}_{١٢-٣٤}$ لمعامل الارتباط الجزئى بين المتغيرين $\text{ص}_١$ ، $\text{ص}_٢$ بعد استبعاد أثر المتغيرين $\text{ص}_٣$ ، $\text{ص}_٤$ من كل منهما ، ونصف هذا المعامل حينئذ بأنه معامل ارتباط جزئى من المرتبة الثانية . ويحسب هذا المعامل من معاملات الارتباط الجزئية من المرتبة الأولى كالآتى :

$$\text{ر}_{١٢-٣٤} = \frac{\text{ر}_{١٢-٤} - \text{ر}_{١٣-٤} \cdot \text{ر}_{٢٣-٤}}{\sqrt{(١ - \text{ر}^{٢}_{١٣-٤})\,(١ - \text{ر}^{٢}_{٢٣-٤})}} \qquad (٢٣)$$

أو كالآتى :

$$\text{ر}_{١٢-٣٤} = \frac{\text{ر}_{١٢-٣} - \text{ر}_{١٤-٣} \cdot \text{ر}_{٢٤-٣}}{\sqrt{(١ - \text{ر}^{٢}_{١٤-٣})\,(١ - \text{ر}^{٢}_{٢٤-٣})}} \qquad (٢٤)$$

وتختبر دلالة هذا المعامل من الصيغة :

$$\text{ت} = \frac{\text{ر}_{21-43}\sqrt{\text{ن} - 4}}{\sqrt{1 - \text{ر}^2_{21-43}}} \qquad \text{بدرجات حرية ن} - 4 \quad (25)$$

ويمتد مفهوم معاملات الارتباط الجزئية لأى عدد منتهى ك من المتغيرات العشوائية التى تشترك فى توزيع معتدل متعدد المتغيرات . ومعامل الارتباط الجزئى من المرتبة م (حيث م < ك − ٢) هو معامل الارتباط بين اثنين من المتغيرات بعد استبعاد أثر م من المتغيرات العشوائية الأخرى . ويعتمد حساب معامل الارتباط الجزئى من المرتبة م على معاملات الارتباط الجزئية من المرتبة السابقة عليها أى من المرتبة م − ١ . هذا ويجدر بنا ملاحظة التماثل فى صيغ هذه المعاملات . أما اختبار دلالة معامل الارتباط الجزئى من المرتبة م فهو تعميم للصيغتين (٢١) ، (٢٥) ويأخذ الصيغة الآتية :

$$\text{ت} = \frac{\text{ر}\sqrt{\text{ن} - 2 - \text{م}}}{\sqrt{1 - \text{ر}^2}} \qquad \text{بدرجات حرية ن} - 2 - \text{م} \quad (26)$$

حيث ر هو معامل الارتباط الجزئى من المرتبة م .

تمارين (١٢ − ٢)

(١) فى عينة عشوائية من ٣٠ وحدة من مجتمع معتدل ذى ثلاثة متغيرات وجد أن معامل الارتباط المتعدد $\text{ر}_{1(32)}$ = ٠,٥ . بيّن أن هذه القيمة تدل على وجود ارتباط فى المجتمع بين المتغير ص_1 والمتغيرين ص_2 ، ص_3 مستخدما مستوى الدلالة ٠,٠٥

(٢) فى عينة عشوائية حجمها ٣٩ من مجتمع معتدل ذى ثلاثة متغيرات وجد أن معاملات الارتباط البسيط بين أزواج هذه المتغيرات هى $ر_{٢١}$ = ٠,٢٨ ، $ر_{٣٢}$ = ٠,٤٩ ، $ر_{٣١}$ = ٠,٥١ اثبت أن $ر_{٢-٣١}$ = ٠,٤٤٥ ، $ر_{١-٢٣}$ = ٠,٤٢ واختبر دلالة كل منهما .

(٣) بين أن معامل الارتباط الجزئى من المرتبة الثانية والذى قيمته $ر_{٤٣-٢١}$ = ٠,٥ المحسوب من عينة عشوائية حجمها ٢٠ من مجتمع معتدل ذى أربعة متغيرات يكون ذا دلالة عند المتسوى ٠,٠٥

(٤) فى عينة عشوائية من ٥٤ من السيدات حسبت المقادير المستهلكة فى فترة ما من نوعين من الغذاء هما البروتين ($ص_{١}$) والدهون ($ص_{٢}$) كما حسب العمر $ص_{٣}$ وقد وجدت معاملات الارتباط البسيط الآتية :

$ر_{٢١}$ = ٠,٥٧٨٤ ، $ر_{٣١}$ = −٠,٤٨٦٥ ، $ر_{٣٢}$ = −٠,٥٢٩٦ أوجد معامل الارتباط الجزئى بين استهلاك البروتين واستهلاك الدهون مستقلا عن أثر العمر ، واختبر دلالته عند المستوى ٠,٠١

الفصل الثالث عشر

دالة التمييز

DISCRIMINANT FUNCTION

يتناول هذا الفصل المشكلة الآتية . نفرض أن باحثا يبحث مجتمعين ا ، ب يشتركان بدرجات متفاوتة فى بعض الخواص التى يمكن قياسها عدديا . إذا كان لدى الباحث عينة عشوائية يعلم أنها من أحد هذين المجتمعين ولكنه لا يعلم ما إذا كانت من المجتمع ا أو من المجتمع ب ، وإذا كان الباحث قد قام بقياس وحدات هذه العينة من حيث ك من تلك الخواص وحصل على القيم العددية $س_1$ ، $س_2$ ، . . . ، $س_ك$ فكيف يستخدم هذه القيم لتحديد المجتمع الذى تنتمى إليه العينة ؟

فمثلا إذا قامت إحدى شركات التنقيب عن البترول بحفر بئر ووجدت منطقة رملية على عمق ٤٠٠٠ قدما فكيف تقرر عن طريق قياس بعض الخواص الجيوفيزائية لعينة مأخوذة من هذه المنطقة ما إذا كانت المنطقة تختزن بترولا (المجتمع أ) فتستمر فى الحفر ، أو لا تختزن بترولا (المجتمع ب) فتتوقف عن الحفر ؟ كذلك إذا كانت إحدى الكليات تقوم بفحص الطلاب المتقدمين إليها فكيف تميز بين الطلاب الذين يصلحون للدراسة فيها (المجتمع ا) والطلاب الذين لا يصلحون لذلك (المجتمع ب) عن طريق إجراء بعض الاختبارات العلمية والنفسية على الطلاب ؟

(١٣ - ١) دالة التمييز :

إن مثل هذه المشكلات تحل عن طريق إيجاد دالة د ($س_1$ ، $س_2$ ، . . . ،

$س_٥$) فى المتغيرات التى تعبر عن تلك الخواص ، وعدد $د_٠$ يقسم قيم هذه الدالة إلى جزءين بحيث إذا كانت قيمة هذه الدالة عند القياسات المشاهدة فى عينة ما أصغر من العدد $د_٠$ تكون العينة من المجتمع أ وإذا كانت أكبر من أو تساوى العدد $د_٠$ تكون العينة من المجتمع ب ، مع بيان احتمال الخطأ فى هذا التقسيم . وتسمى هذه الدالة حينئذ بدالة التمييز كما يسمى العدد $د_٠$ بالنقطة الحدية boundary point .

فمثلا فى حفر بئر البترول قد تكون دالة التمييز هى :

$$د (س_١ ، س_٢ ، ... ، س_٥) = ٢ س_١ - س_٢ + ٣ س_٣ - ٢ س_٤ + ٠,٥ س_٥$$

والعدد $د_٠$ = ٢٦ بحيث إذا كانت قيمة هذه الدالة لقياسات أخذت على وحدات عينة ما أصغر من ٢٦ يستمر الحفر (المجتمع أ) وإذا كانت أكبر من أو تساوى ٢٦ يتوقف الحفر (المجتمع ب) . نفرض مثلا أن القيم المشاهدة للخواص الخمسة فى إحدى العينات هى على الترتيب ٣٠ ، ١٢٠ ، ١٠ ، ١٦ ، ٢٠٠ . بالتعويض بهذه القيم فى دالة التمييز نجد أن د (٣٠ ، ١٢٠ ، ١٠ ، ١٦ ، ٢٠٠) = ٣٨ وبما أن ٣٨ أكبر من ٢٦ تكون العينة من المجتمع ب ويتوقف الحفر .

(١٣ - ٢) إيجاد دالة التمييز :

كيف نعد الدالة د والعدد $د_٠$ لتكون جاهزة للتطبيق على عينات يراد معرفة المجتمع الذى تنتمى إليه ؟ الطريق إلى ذلك ما يلى :

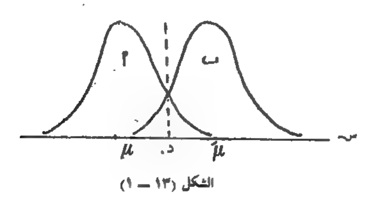
نأخذ عينة عشوائية حجمها $ن_١$ من المجتمع الأول وعينة عشوائية حجمها $ن_٢$ من المجتمع الثانى ، ونحدد ك من الخواص التى نرى أنها ذات تأثير فى التمييز بين المجتمعين (ك < $ن_١$ ، ك < $ن_٢$) ثم نقيس كل وحدة من وحدات العينتين من حيث هذه الخواص فتكون القياسات الناتجة وعددها ($ن_١$ + $ن_٢$) ك هى الأساس الذى نبنى عليه دالة التمييز كما سيتبين بعد . فمثلا إذا كانت ك = ٤ تكون البيانات كما فى الجدول (١٣ - ١) الآتى :

الجدول (١٣ – ١)

	عينة المجتمع الأول					عينة المجتمع الثانى			
	$س_{١}$	$س_{٢}$	$س_{٣}$	$س_{٤}$		$س'_{١}$	$س'_{٢}$	$س'_{٣}$	$س'_{٤}$
١	$س_{١١}$	$س_{٢١}$	$س_{٣١}$	$س_{٤١}$	١	$س'_{١١}$	$س'_{٢١}$	$س'_{٣١}$	$س'_{٤١}$
٢	$س_{١٢}$	$س_{٢٢}$	$س_{٣٢}$	$س_{٤٢}$	٢	$س'_{١٢}$	$س'_{٢٢}$	$س'_{٣٢}$	$س'_{٤٢}$
..	..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..
$ن_{١}$	$س_{١،ن}$	$س_{٢،ن}$	$س_{٣،ن}$	$س_{٤،ن}$	$ن_{٢}$	$س'_{١،ن}$	$س'_{٢،ن}$	$س'_{٣،ن}$	$س'_{٤،ن}$
المتوسط	$\bar{س}_{١}$	$\bar{س}_{٢}$	$\bar{س}_{٣}$	$\bar{س}_{٤}$	المتوسط	$\bar{س}'_{١}$	$\bar{س}'_{٢}$	$\bar{س}'_{٣}$	$\bar{س}'_{٤}$

(أولا) : حالة متغير واحد :

لتقديم الفكرة التى تؤسس عليها دالة التمييز ، نبدأ بالحالة التى يكون لدينا فيها متغير واحد س (ك = ١) . سنفترض أن لهذا المتغير توزيعا معتدلا بمتوسط μ للمجتمع الأول ، μ' للمجتمع الثانى ، أما التباين فقيمته واحدة للمجتمعين وقدرها σ^2 . يمثل هذان التوزيعان كما فى الشكل (١٣ – ١) الآتى .



الشكل (١٣ – ١)

نفرض أن س هى قيمة المتغير س التى وجدناها فى عينة نعلم أنها من أحد المجتمعين ونريد تحديد المجتمع الذى تنتمى إليه العينة . إن دالة التمييز هنا تكون على الصورة د (س) = س . من الطبيعى أن نأخذ متوسط المجتمعين د = $\frac{1}{2}(\mu + \mu')$ كنقطة حدية تستخدم للفصل بينهما . وإذا فرضنا أن μ أصغر من μ' فإن قاعدة التمييز تكون كالآتى :

إذا كان س $< \frac{1}{2}(\mu + \mu')$ نقرر أن العينة من المجتمع ا

وإذا كان س $\geqslant \frac{1}{2}(\mu + \mu')$ نقرر أن العينة من المجتمع ب

ونتساءل الآن متى يكون قرارنا خاطئا ؟ إن هذا الخطأ يقع حين تكون العينة فى الواقع من المجتمع ا ويكون قرارنا أنها من المجتمع ب وهذا يحدث حينما

س $\geqslant \frac{1}{2}(\mu + \mu')$ (مع أن العينة من المجتمع أ)

أى حينما $\frac{\text{س} - \mu}{\sigma} \geqslant \frac{\frac{1}{2}(\mu + \mu') - \mu}{\sigma} = \frac{\mu' - \mu}{2\,\sigma}$

أى حينما ع $\geqslant \frac{\delta}{2\,\sigma}$

حيث ع = $\frac{\text{س} - \mu}{\sigma}$ ، $\delta = \mu' - \mu$ (١)

واحتمال هذا الخطأ هو ل (ع $\geqslant \frac{\delta}{2\,\sigma}$) (٢)

ويسمى هذا الاحتمال باحتمال خطأ التقسيم . ونحصل على نفس قيمة هذا الاحتمال حين تكون العينة من المجتمع ب ونقرر أنها من المجتمع ا . تحقق من ذلك .

ولما كانت ع تتبع التوزيع المعتدل القياسى فإننا نستطيع إيجاد هذا الاحتمال من جدول المساحات أسفل المنحنى المعتدل القياسى .

ويلاحظ أنه إذا كان المتوسطان μ ، μ' متساويين أو متساويين تقريبا فإن $\frac{\delta}{2\sigma}$ تكون تقريبا مساوية للصفر ويكون احتمال خطأ التقسيم مساويا بالتقريب للنسبة ٥٠٪ ومعنى هذا أن التقسيم عشوائيا وفى هذه الحالة لا يكون هناك جدوى من إيجاد دالة صالحة للتمييز بين المجتمعين . كما يلاحظ أن احتمال خطأ التقسيم يكون أصغر ما يمكن إذا كانت $\left|\frac{\delta}{\sigma}\right| = \left|\frac{\mu - \mu'}{\sigma}\right|$ أكبر ما يمكن أى إذا كانت القيمة الموجبة لنسبة الفرق بين متوسطى المجتمعين إلى الانحراف المعيارى المشترك أكبر ما يمكن ، وتستخدم هذه الحقيقة كأساس لاشتقاق أفضل دالة للتمييز بين المجتمعين .

(ثانيا) : حالة ك من المتغيرات :

نأتى الآن للحالة التى يكون لدينا فيها ك من المتغيرات $س_١$ ، $س_٢$ ، ... ، $س_ك$. كما سبق القول ، للتوصل إلى دالة التمييز نأخذ عينة عشوائية من كلا المجتمعين ونقيس قيم هذه المتغيرات لكل وحدة من وحدات العينتين فتكون البيانات الناتجة هى الأساس الذى نبنى عليه إنشاء أفضل دالة تمييز كما يتبين بعد .

لإمكانية التحليل الاحصائى سنضع الافتراضات الثلاثة الآتية .

افتراضات التحليل :

(١) دالة التمييز خطية أى على الصورة

$$د(س) = \alpha_١ س_١ + \alpha_٢ س_٢ + \dots + \alpha_ك س_ك$$

$$= \sum_{ر=١}^{ك} \alpha_ر س_ر \qquad (٣)$$

حيث α_1 ، α_2 ، .... ، α_k بارامترات مجهولة مطلوب تقديرها من العينتين بشرط أن نحصل من هذه التقديرات على أقل احتمال لخطأ التقسيم .

(٢) المتغيرات س١ ، س٢ ، ... ، سك هى متغيرات عشوائية توزيعها المشترك هو توزيع معتدل متعدد (ذو ك من المتغيرات) سواء فى المجتمع ا أو فى المجتمع ب . وتأخذ متوسطات هذه المتغيرات القيم μ_1 ، μ_2 ، ... ، μ_k فى المجتمع الأول ، والقيم μ'_1 ، μ'_2 ، ... ، μ'_k فى المجتمع الثانى . أما التباين σ^2_r لأى متغير سر والذى سنرمز له بالرمز σ_{rr} فهو واحد فى المجتمعين وتختلف التباينات من متغير إلى آخر ، كما أن التغاير σ_{rs} (ر ≠ و) لمتغيرين سر ، سو يكون واحدا فى المجتمعين وتختلف التغايرات بين كل زوج وآخر من المتغيرات .

من هذا الافتراض ينتج أن دالة التمييز المعرفة فى (٣) تكون ذات توزيع معتدل لأنها خطية فى متغيرات معتدلة . هذا مع تذكر أن اعتدالية التوزيع المتعدد المشترك تتضمن اعتدالية كل متغير على حدة .

(٣) العينات التى تؤخذ من كل مجتمع هى عينات عشوائية .

لايجاد أفضل تقدير لدالة التمييز المعرفة فى (٣) من أزواج العينات ينبغى أن نقدر البارامترات α_1 ، α_2 ، ... ، α_k بأعداد ا١ ، ا٢ ، .... ، اك تجعل الدالة مجـ ار سر ذات أقل خطأ ممكن للتقسيم . وهذا الشرط كما جاء فى حالة المتغير الواحد يكافىء أن تكون النسبة $\left|\frac{\delta}{\sigma}\right|$ أكبر ما يمكن حيث δ ، σ هنا تعرفان كالآتى :

δ = الفرق الكلى بين المتوسطات المتناظرة فى المجتمعين مرجحة بالمعاملات α_r

$$= \sum_{1}^{k} \alpha_r (\mu_r - \mu'_r)$$

$$= \sum_{1}^{k} \alpha_r \delta_r \qquad \text{حيث } \delta_r = \mu_r - \mu'_r \qquad (٤)$$

، σ^2 = التباين الكلى للدالة (٣) وهى مجـ $\alpha_{ر}$ س$_{ر}$

$$= \sum_{ر=1}^{ك} \sum_{ن=1}^{ك} \alpha_{ر}\, \alpha_{ن}\, \sigma_{رن} \qquad (٥)$$

فمثلا إذا كانت ك = ٣ فإن

$$\delta = \alpha_1(\mu_1 - \bar{\mu}_1) + \alpha_2(\mu_2 - \bar{\mu}_2) + \alpha_3(\mu_3 - \bar{\mu}_3)$$

$$، \sigma^2 = \alpha_1^2 \sigma_{11} + \alpha_2^2 \sigma_{22} + \alpha_3^2 \sigma_{33} + 2\alpha_1 \alpha_2 \sigma_{12}$$
$$+ 2\alpha_1 \alpha_3 \sigma_{13} + 2\alpha_2 \alpha_3 \sigma_{23}$$

المطلوب إذن إيجاد التقديرات $\alpha_{ر}$ بحيث تكون القيمة المطلقة $\Delta = \left| \frac{\delta}{\sigma} \right|$

أكبر ما يمكن حيث δ ، σ معرفتان فى (٤) ، (٥) . ولتجنب الإشارات سنأخذ $\Delta^2 = \frac{\delta^2}{\sigma^2}$ بدلا من Δ . أى أننا سنقدر البارامترات بالقيم التى تجعل الدالة الآتية أكبر ما يمكن :

$$\Delta^2 = \frac{\delta^2}{\sigma^2} = (\text{مجـ}\ \alpha_{ر}\, \delta_{ر})^2 \Big/ \text{مجـ مجـ}\ \alpha_{ر}\, \alpha_{ن}\, \sigma_{رن}$$

$$= (\text{مجـ}\ \alpha_{ر}\, \delta_{ر})^2 \Big/ \frac{1}{ن_1 + ن_2 - 2}\ \text{مجـ مجـ}\ \alpha_{ر}\, \alpha_{ن}\, م_{رن} \qquad (٦)$$

حيث $\text{م}_{\text{رر}}$ = مجموع المربعات للمتغير $\text{س}_{\text{ر}}$ فى المجتمعين (ر = ١ ، ٢ ، . . . ، ك)

، $\text{م}_{\text{رى}}$ = مجموع حواصل الضرب للمتغيرين $\text{س}_{\text{ر}}$ ، $\text{س}_{\text{ى}}$ فى المجتمعين .

وتعظيم الدالة (٦) يكافى تعظيم الدالة

$$\left(\sum \alpha_{\text{ر}}\, \delta_{\text{ر}}\right)^{2} \Big/ \sum\sum \alpha_{\text{ر}}\, \alpha_{\text{ى}}\, \text{م}_{\text{رى}} \qquad (٧)$$

التى لا تختلف عنها إلا فى المقدار الثابت $\text{ن}_{1} + \text{ن}_{2} - 2$. ويقتضى هذا التعظيم إجراء التفاضل الجزئى لهذه الدالة بالنسبة إلى كل من α_{1} ، α_{2} ، ... ، $\alpha_{\text{ك}}$ والمساواة بالصفر . وهذا يؤدى إلى الحصول على ك من المعادلات الخطية فى ك من المجاهيل تشبه المعادلات المعتادة فى الانحدار الخطى ، فهى تأخذ الصورة الآتية :

$$\text{م}_{11}\alpha_{1} + \text{م}_{21}\alpha_{2} + \ldots + \text{م}_{\text{ك}1}\alpha_{\text{ك}} = \delta_{1}$$

$$\text{م}_{12}\alpha_{1} + \text{م}_{22}\alpha_{2} + \ldots + \text{م}_{\text{ك}2}\alpha_{\text{ك}} = \delta_{2}$$

$$\ldots\ldots \quad \ldots\ldots \quad \ldots\ldots \quad \ldots\ldots \quad \ldots\ldots \quad \ldots\ldots \qquad (٨)$$

$$\text{م}_{1\text{ك}}\alpha_{1} + \text{م}_{2\text{ك}}\alpha_{2} + \ldots + \text{م}_{\text{كك}}\alpha_{\text{ك}} = \delta_{\text{ك}}$$

والقيم ا_{1} ، ا_{2} ، . . . ، $\text{ا}_{\text{ك}}$ التى تحقق هذه المعادلات جميعا تكون هى التقديرات المطلوبة للبارامترات α_{1} ، α_{2} ، ... ، $\alpha_{\text{ك}}$ وبالتالى تكون دالة التمييز هى :

$$\text{ا}_{1}\,\text{س}_{1} + \text{ا}_{2}\,\text{س}_{2} + \ldots + \text{ا}_{\text{ك}}\,\text{س}_{\text{ك}} \qquad (٩)$$

هذا مع ملاحظة أن مجاميع المربعات ومجاميع حواصل الضرب تقدر من العينتين كالآتى :

$\text{م}_{١١}$ = مجموع المربعات لقيم المتغير $\text{س}_{١}$ فى العينة الأولى + مجموع المربعات لقيم نفس المتغير فى العينة الثانية .

وبالمثل لمجاميع المربعات الأخرى $\text{م}_{٢٢}$ ، $\text{م}_{٣٣}$ ، . . . ، $\text{م}_{\text{كك}}$

$\text{م}_{٢١}$ = مجموع حواصل الضرب للمتغيرين $\text{س}_{١}$ ، $\text{س}_{٢}$ فى العينة الأولى + مجموع حواصل الضرب لنفس المتغيرين فى العينة الثانية

وبالمثل لمجاميع حواصل الضرب الأخرى $\text{م}_{\text{رو}}$ $(\text{ر} \neq \text{و})$

أما التباينات والتغايرات فنحصل عليها بالقسمة على درجات الحرية وهى $\text{ن}_{١} + \text{ن}_{٢} - ٢$.

وإيجاد هذه القيم ثم حل المعادلات (٨) يحتاج إلى الحاسب الالكترونى توفيرا للجهد وضمانا للدقة والسرعة .

النقطة الحدية :

كما فى حالة المتغير الواحد ، من الطبيعى أن نأخذ متوسط المجتمعين كنقطة حدية تفصل بينهما ، وعلى ذلك تقدر النقطة الحدية $\text{د}_{\text{.}}$ من العينتين كالآتى :

$$\text{د}_{\text{.}} = \frac{١}{٢} \sum_{١}^{\text{ك}} \text{ا}_{\text{ر}} \left(\bar{\text{س}}_{\text{ر}} + \bar{\text{س}}'_{\text{ر}}\right) \qquad (١٠)$$

وهكذا نكون قد حصلنا على دالة التمييز (٩) والقيمة الحدية (١٠) . فإذا حصلنا على قياسات $\text{س}_{١}$ ، $\text{س}_{٢}$ ، . . . ، $\text{س}_{\text{ك}}$ لعينة جديدة ، يمكن بالتعويض بهذه القياسات فى الدالة (٩) ثم المقارنة بالعدد $\text{د}_{\text{.}}$ أن نقرر ما إذا كانت العينة تنتمى إلى المجتمع ا أو إلى المجتمع ب .

احتمال خطأ التقسيم PROBABILITY OF MISCLASSIFICATION

يبقى أن نقدر إحتمال خطأ التقسيم لدالة التمييز التى أوجدناها . وكما فى حالة المتغير الواحد تكون أكبر قيمة لهذا الاحتمال هى المعطاة بالصيغة (٢) وهى :

$$ل\ (ع \geqslant \frac{\delta}{٢\ \sigma}) \qquad (١١)$$

حيث δ ، σ تعطيان بالصيغتين (٤) ، (٥)، ويمكن إثبات أن $\frac{\delta}{\sigma}$ تقدر من العينتين كما يلى :

$$تقدير\ \frac{\delta}{\sigma} = \sqrt{(ن_١ + ن_٢ - ٢) \sum^{ك} ا_ر ف_ر} \qquad حيث\ ف_ر = \bar{س}_ر - \bar{س}'_ر \quad (١٢)$$

، $\bar{س}_ر$ الوسط الحسابى للمتغير $س_ر$ فى العينة الأولى

، $\bar{س}'_ر$ الوسط الحسابى للمتغير $س_ر$ فى العينة الثانية .

والاحتمال (١١) يمكن إيجاده من جدول المساحات أسفل المنحنى المعتدل القياسى وهو يشير إلى درجة الثقة فى دالة التمييز ، ويعتبر التقسيم جيدا بدرجة عالية إذا لم يزد هذا الاحتمال عن ٧٪ .

ملاحظات :

(١) إذا كانت دالة التمييز هى د ($س_١$ ، $س_٢$ ، ... ، $س_ك$) والقيمة الحدية د. ، وحسبنا من أزواج العينات القيمتين د ($\bar{س}_١$ ، $\bar{س}_٢$ ، ... ، $\bar{س}_ك$) ،

د ($\bar{س}_1$ ، $\bar{س}_2$ ، ... ، $\bar{س}_ن$) فينبغي أن تكون إحدى هاتين القيمتين أكبر من د. والأخرى أصغر منها .

(٢) إذا كان كل زوج من متوسطات المجتمعين متساويين ($\mu_ر = \mu'_ر$) أو متساويين تقريبا فلا أمل فى العثور على دالة تميز بين المجتمعين بكفاءة . ولهذا تجب العناية باختيار المتغيرات التى تستخدم فى دوال التمييز بحيث تكون هناك فروق معقولة بين متوسطاتها .

(٣) استخدام الحاسب الآلى فى هذه الدراسة أمر ضرورى لسرعة ودقة ما نريد التوصل إليه ، خاصة وأننا نضطر فى كثير من الحالات إلى تجربة مجموعات مختلفة من المتغيرات لاختيار الأصلح منها . وهذا يتطلب مشقة كبيرة يغنينا عنها الحاسب .

مثال (١٣ – ١) :

فى دراسة لتوزيع بكتريا النيتروجين Azotobacter فى التربة كان المطلوب معرفة مدى دقة التنبؤ بوجود أو عدم وجود هذه البكتريا فى التربة ، أو بمعنى آخر المطلوب إيجاد دالة خطية تميز التربة التى تحتوى على هذه البكتريا من التربة التى لا تحتوى عليها ، وذلك باستخدام ثلاث خواص كيميائية للتربة هى :

$س_1$ = الأس الأيدروجينى pH ، $س_2$ = كمية الفوسفات الموجودة ، $س_3$ = المحتوى الكلى للنيتروجين . وقد أخذت عينة عشوائية حجمها $ن_1$ = ١٠٠ من تربة لا تحتوى على هذه البكتريا وعينة عشوائية حجمها $ن_2$ = ١٨٦ من تربة تحتوى عليها . وقيست كل وحدة من وحدات العينتين من حيث هذه الخواص الثلاث وحسبت الفروق $ف_ر = \bar{س}_ر - \bar{س}'_ر$ بين أزواج المتوسطات فى العينتين فوجدت كما يلى :

$ف_1$ = ٠,١٤٠٨ ، $ف_2$ = ٠,٠٨٢١ ، $ف_3$ = ٠,٠٨٢٦

وحسبت مجاميع المربعات $م_{سس}$ لكل من المتغيرات الثلاثة ، ومجاميع حواصل الضرب $م_{رى}$ لأزواج المتغيرات فوجدت كما يلى :

$$م_{١١} = ١,١١١ ، م_{٢٢} = ١,٠٤٣ ، م_{٣٣} = ٢,٩٤٢$$

$$، م_{٢١} = ٠,٢٢٩ ، م_{٣١} = ٠,١٩٨ ، م_{٣٢} = ٠,٠٥١$$

وبذلك كانت المعادلات المعتادة كالآتى :

$$١,١١١ \alpha_{١} + ٠,٢٢٩ \alpha_{٢} + ٠,١٩٨ \alpha_{٣} = ٠,١٤٠٨$$

$$٠,٢٢٩ \alpha_{١} + ١,٠٤٣ \alpha_{٢} + ٠,٠٥١ \alpha_{٣} = ٠,٠٨٢١$$

$$٠,١٩٨ \alpha_{١} + ٠,٠٥١ \alpha_{٢} + ٢,٩٤٢ \alpha_{٣} = ٠,٠٨٢٦$$

استخدم الحاسب الالكترونى فى حل هذه المعادلات فأنتج القيم الآتية :

$ا_{١} = ٠,١١٢٢٩$ ، $ا_{٢} = ٠,٠٥٣١٠$ ، $ا_{٣} = ٠,٠١٩٦٠$ وبذلك كانت دالة التمييز هى :

$$د(س_{١} ، س_{٢} ، س_{٣}) = ٠,١١٢٢٩ س_{١} + ٠,٠٥٣١٠ س_{٢} + ٠,٠١٩٦٠ س_{٣}$$

ويمكن حساب النقطة الحدية $د_{.}$ من الصيغة (١٠) .

نقدر أكبر احتمال لخطأ التقسيم بهذه الدالة كما يلى :

$$ع ا_{ر} ف_{ر} = ٠,١١٢٢٩ \times ٠,١٤٠٨ + ٠,٠٥٣١٠ \times ٠,٠٨٢١ + ٠,٠١٩٦ \; ٠,٠٨٢٢٦$$
$$= ٠,٠٢١٧٩$$

من (١٢) : قيمة $\frac{\delta}{\sigma}$ مقدرة من العينة = $\sqrt{(١٠٠ + ١٨٦ - ٢) \times ٠,٠٢١٧٩}$

= ٢,٤٨٧٦

من (١١) أكبر احتمال لخطأ التقسيم = ل (ع ≥ $\frac{\delta}{٢\sigma}$)

= ل (ع ≥ ١,٢٤٣٨)

= ١٠,٧٥٪ (من جدول المساحات أسفل المنحنى المعتدل القياسى)

قد يكون من المناسب مقارنة هذا الاحتمال باحتمال خطأ التقسيم الناشىء من دالة تمييز استخدمت واحدا فقط من المتغيرات الثلاث .

نفرض أننا استخدمنا المتغير سـ$_١$ فقط . لدينا :

ف$_١$ = ٠,١٤٠٨ وهذه تقدير لقيمة δ

الانحراف المعيارى = $\sqrt{\frac{١}{ن_١ + ن_٢ - ٢}\, م_{١١}} = \sqrt{\frac{١,١١١}{٢٨٤}} = ٠,٠٦٢٥$

وهذا تقدير σ .

∴ ل (ع ≥ $\frac{\delta}{٢\sigma}$) = ل(ع ≥ $\frac{٠,١٤٠٨}{٢ \times ٠,٠٦٢٥}$) = ل(ع ≥ ١,١٢٦)

= ١٢,٩٢٪

وإذا استخدمنا المتغير س_{2} فقط نجد بنفس الطريقة ما يلى :

$$\text{ل}\ (\text{ع} \geqslant \frac{\delta}{\text{٢}\,\sigma}) = \text{ل}\ (\text{ع} > \text{٠,٥٨}) = \text{٢٨,١}\%$$

أما إذا استخدمنا المتغير س_{3} فقط فإن

$$\text{ل}\ (\text{ع} \geqslant \frac{\delta}{\text{٢}\,\sigma}) = \text{ل}\ (\text{ع} \geqslant \text{٠,٤١}) = \text{٣٤,٠٩}\%$$

ويتضح أن أفضل دالة تمييز تؤسس على متغير واحد فقط هى تلك التى تستخدم المتغير س_{1} ويليها تلك التى تستخدم المتغير س_{2} ثم تلك التى تستخدم المتغير س_{3} وهذه الأخيرة تكون دالة ضعيفة للغاية لا يجوز الاعتماد عليها . على أن دالة التمييز المركبة من المتغيرات الثلاثة معا هى أفضلها جميعا .

(١٣ – ٣) اختبار تساوى أزواج المتوسطات – اختبار ت^{2} .

كما سبق الذكر فى الملاحظة (٢) بالبند السابق ينبغى أن تكون المتغيرات التى تستخدم فى إعداد دالة التمييز هى تلك المتغيرات الأكثر قدرة على التمييز بين المجتمعين أى التى تختلف متوسطاتها فى المجتمعين اختلافا معقولا . ولذلك يهمنا فى اختيار هذه المتغيرات أن نبحث دلالة الفروق بين أزواج المتوسطات فى العينتين ، فإذا كانت هذه الفروق ليست ذات دلالة بمعنى أن أزواج المتوسطات فى المجتمعين متساوية ، لا تكون دالة التمييز قادرة على فصل المجتمعين . أما إذا كانت تلك الفروق جوهرية فإن فرصة دالة التمييز فى تقسيم المجتمعين بكفاءة تكون كبيرة .

وهناك اختبار ابتكره هوتلنج Hotelling يسمى اختبار ت^{2} يمكننا من اختبار

الفرض الصفرى المركب عن تساوى كل زوج من المتوسطات ، أى اختبار الفرض الصفرى :

$\mu_{ر} = \mu_{ر}$ لجميع ر = ١ ، ٢ ، ... ، ك (١٣)

ويعتمد هذا الاختبار على تحليل التباين للمتغير مج ار سر إلى مركبتين : « بين » و« داخل » المجتمعين كالآتى :

$$م^{٢} \text{(داخل العينات)} = \sum_{ر=١}^{ك} \sum_{و=١}^{ن} ا_{ر}\, ا_{و}\, م_{رو}$$

$$= \text{مج } ا_{ر}\, ف_{ر} \text{ بدرجات حرية } ن_{١} + ن_{٢} - ك - ١ \quad (١٤)$$

(تنتج الصيغة الأخيرة من ضرب المعادلات (٨) فى $\alpha_{١}$ ، $\alpha_{٢}$ ، ... ، $\alpha_{ك}$ ثم الجمع مع وضع $ا_{ر}$ بدلا من $\alpha_{ر}$ ، $ف_{ر}$ بدلا من $\delta_{ر}$) .

$$م^{٢} \text{(بين العينات)} = \frac{ن_{١}\, ن_{٢}}{ن_{١} + ن_{٢}} (\text{مج } ا_{ر}\, ف_{ر})^{٢} \text{ بدرجات حرية ك} \quad (١٥)$$

وقد أخذت ك كدرجة الحرية بين العينات لأن التقديرات $ا_{ر}$ اختيرت على أساس تعظيم النسبة بين $م^{٢}$ (بين العينات) ، $م^{٢}$ (داخل العينات) .

فى المثال السابق نجد ما يلى :

$$م^{٢} \text{(داخل العينات)} = \text{مج } ا_{ر}\, ف_{ر}$$

= ٠,٠٢١٧٩ (سبق حسابه) بدرجات حرية ٢٨٢

م م (بين العينات) $= \frac{ن_1 ن_2}{ن_1 + ن_2} (مج ا_ر ف_ر)^2$

$= \frac{١٠٠ \times ١٨٦}{٢٨٦} (٠,٠٢١٧٩)^2$

$= ٠,٠٣٠٨٨$ بدرجات حرية ٣

وينشأ لدينا جدول التباين الآتى .

الجدول (١٣ – ٢)

تحليل التباين لدالة التمييز – اختبار ت²

مصدر التباين	م م	درجات الحرية	ط م
بين نوعى التربة	$\frac{ن_1 ن_2}{ن_1 + ن_2}(مج ا_ر ف_ر)^2 = ٠,٠٣٠٨٨$	ك = ٣	٠,٠١٠٢٩
داخل نوعى التربة	مج ا_ر ف_ر = ٠,٠٢١٧٩	ن_1 + ن_2 – ك – ١ = ٢٨٢	٠,٠٠٠٠٧٧

$ف_ى = ٠,٠١٠٢٩ \div ٠,٠٠٠٠٧٧ = ١٣٣,١^{**}$

هذه القيمة أكبر بكثير من القيمة الحرجة $ف_{٠,٠١ [٣، ٢٨٢]}$ التى لا تزيد عن ٣,٧٨ فهى ذات دلالة عالية وتدعونا إلى رفض الفرض الصفرى عن تساوى أزواج المتوسطات فى المجتمعين ، وهذا ما يجب أن يكون إذا كانت دالة التمييز ذات كفاءة فى فصل المجتمعين .

يلاحظ أنه إذا ظهر أن ف غير ذات دلالة فإن دالة التمييز تفشل فى فصل المجتمعين وينبغى حينئذ البحث عن متغيرات أخرى أو البحث عن دالة أخرى غير خطية فيما لدينا من متغيرات .

(١٣ – ٤) استخدامات دالة التمييز :

إن دالة التمييز هى وسيلة لدراسة مدى تداخل المجتمعات فى بعضها أو مدى تباعدها عن بعضها . ولهذه الدالة ثلاثة أنواع من الاستخدامات تتلخص فيما يلى :

(١) التقسيم والتشخيص :

يتبين هذا الهدف من المثال المقدم فى بداية هذا الفصل حيث استخدمت دالة للتمييز بين المناطق التى تحتوى بترولا والمناطق التى لا تحتوى عليه مستعينة بخمسة متغيرات جيوفيزيائية ، وكذلك من المثال (١٣ – ١) حيث أعدت دالة فى ثلاثة متغيرات كيميائية تقسم التربة إلى نوعين يحتوى أحدهما على بكتريا النيتروجين ولا يحتوى الآخر عليها . كذلك إذا كان هناك نوعان من الحمى يتشابهان فى الأعراض فمن المفيد أن يكتشف الطبيب القياسات الجسمية والمعملية التى تساعده على التمييز بين نوعى الحمى وأن يعرف الطريقة المثلى لضم هذه القياسات فى دالة واحدة ، وكيفية تقدير درجة الثقة فى تشخيص المرض .

(٢) دراسة العلاقات بين المجتمعات :

فمثلا ، إلى أى مدى تختلف الاتجاهات والاستعدادات النفسية للمهندس الكفء عنها فى رجل الأعمال الكفء ؟ أو هل يختلف المدخنون عن غير المدخنين اختلافا جوهريا فى السمات النفسية والعادات السلوكية ؟

(٣) تعميم لاختبار ت :

إذا أجرينا عدة قياسات على كل من عينتين عشوائيتين أخذتا من مجتمعين معتدلين ورغبنا فى استخدام اختبار واحد للفرض الصفرى عن تساوى متوسطات

هذين المجتمعين بالنسبة لجميع هذه القياسات فإن دالة التمييز تمكننا من ذلك عن طريق اختبار ت[٢] كما جاء بالبند السابق .

هذا وتجدر الإشارة إلى أن الدراسة التى قدمت فى هذا الفصل عن دالة التمييز تمتد إلى الحالات التى تتناول أكثر من مجتمعين ، كما تمتد إلى الحالات التى تكون فيها دالة التمييز غير خطية .

الفصل الرابع عشر

الطرق غير البارامترية

NONPARAMETRIC METHODS

إن معظم اختبارات الفروض التي جاءت في الفصول السابقة كانت تتطلب افتراض أن للمجتمع الذى نعاين منه توزيعاً معتدلا أو يمكن تقريبه بتوزيع معتدل ، كما أن بعضها كان يتطلب افتراضات أخرى مثل تساوى التباينات أو استقلال العينات . وجدير بالذكر أن هذه الاختبارات يمكن الاطمئنان إليها حتي لو وجدت انحرافات بسيطة عن هذه الافتراضات غير أن هناك مواقف يستحيل فيها قبول مثل هذه الافتراضات ولهذا كان من الضرورى أن تنشأ أساليب أخرى تبني على افتراضات أقل صرامة . وقد عرفت هذه الأساليب بالطرق غير البارامترية أو بالطرق حرة التوزيع distribution-free methods وهى طرق يمكن تطبيقها على مدى واسع من التوزيعات دون أن تفترض توزيعاً محدداً لما تتناوله من مجتمعات . وتوصف هذه الطرق بأنها غير بارامترية لأن أغلبها لا يهتم باختبار أو تقدير بارامترات (أدلة) المجتمع .

وفضلا عن أن الطرق والاختبارات غير البارامترية تستخدم تحت شروط عامة للغاية وتعفيناً من القلق عن صحة الافتراضات فهى تتميز بعدة أمور منها :

(١) أنها عادة ما تكون أسهل في الفهم والتفسير من تلك الطرق القياسية التي تعمل كبديل لها .

(٢) أن ما تتطلبه من عمليات حسابية تكون عادة سهلة وسريعة .
(٣) أنها لا تشترط أن تكون البيانات كمية (عددية) بل يمكن أن تكون نوعية أو ترتيبية .

ولهذا شاع استخدام الطرق غير البارامترية بالرغم من أنها لا تعطى نفس القدر من المعلومات أو الدقة التي تعطيها الطرق البارامترية المناظرة لها فهى بصفة عامة أقل كفاءة منها . وإذا وجد موقف يمكن فيه تطبيق كلا الأسلوبين فينبغى دائماً استخدام الأسلوب البارامترى فهو الأكثر كفاءة .

وقد مر بنا مثالان للاختبارات غير البارامترية جاء أحدهما بالبند (٦ — ٧) عند استخدام اختبار χ^2 ، وجاء الآخر في البندين (١٠ — ٧) و(١٠ — ٨) عند دراسة معامل ارتباط الرتب (سبيرمان) ودلالته ، ونقدم فيما يلى ستة من الاختبارات الأخرى الشهيرة .

(١٤ — ١) اختبار التلاحقات (للكشف عن عشوائية العينة) : RUNS TEST

إن جميع طرق الاستدلال الإحصائي التي نوقشت في الفصول السابقة بنيت على افتراض أن العينات عشوائية وعلى أن التجارب التي نحصل منها على البيانات صممت على هذا الأساس . غير أن هناك حالات يصعب فيها تحديد مدى تحقق هذا الافتراض وينبغى حينئذ اختبار عشوائية العينة قبل التصدى لتحليل البيانات .

وتظهر هذه الحالات بصفة خاصة عندما نكون عاجزين كلياً أو جزئياً عن التحكم في اختيار العينة . فمثلا في تقدير معدل الوفاة من مرض معين لا مفر من الاعتماد على سجلات سابقة وهذه لا تشكل عينة عشوائية بالمعنى الدقيق . كذلك الحال حين لا يكون لنا خيار إلا الاعتماد على أى سجلات متاحة لإعطاء تنبؤات عن الأحوال الجوية أو دراسة حوادث المرور أو حين نأخذ وحدات صناعية بحسب ترتيب إنتاجها .

ويعتمد اختبار عشوائية العينة الذى نقدمه هنا على نظرية تسمى بنظرية التلاحقات theory of runs التي تعتمد بدورها على الترتيب الذى سحبت به عناصر العينة . ويهدف الاختبار إلى الكشف عما إذا كان هذا الترتيب عشوائيا أو يتخذ نمطاً معيناً لا يمكن أن نعزوه إلى الصدفة .

اعتبر متتابعة تنقسم عناصرها إلى صنفين فقط . إن أى متتابعة جزئية تتألف من حد أقصي لعناصر متتالية من أحد هذين الصنفين تسمى **تلاحقة** . فمثلا إذا اخترنا ١٢ شخصاً وكان الحرف «ك» يرمز إلى أن الشخص « ذكر » والحرف «ث» يرمز إلى أن الشخص أنثي فإن المتتابعة .

ك ك   ث ث ث   ك   ث ث   ك ك ك ك

تشتمل على ٥ تلاحقات ، تتألف الأولى من اثنين من الكافات ونقول إن طولها اثنان، وتتألف الثانية من ثلاثة من الثاءات ونقول إن طولها ثلاثة ، وتتألف الثالثة من كاف واحدة .. .. وهكذا .

وسواء كانت البيانات نوعية أو كمية فإن اختبار التلاحقات يقسمها إلى صنفين متنافيين : ذكور أو إناث – وحدة معيبة أو غير معيبة – مريض أو غير مريض – فوق الوسيط أو تحت الوسيط .. وهكذا .

ليكن ن = حجم العينة ، س = عدد التلاحقات

، ن$_{1}$ = عدد المرات التي يظهر فيها الحرف الذى يرمز إلى أحد الصنفين .

، ن$_{2}$ = عدد المرات التي يظهر فيها الحرف الذى يرمز إلى الصنف الآخر .

فمثلا في متتابعة التلاحقات المذكورة : لدينا ن = ١٢ ، س = ٥ ، ن$_{1}$ = ٧ ، ن$_{2}$ = ٥

إن اختبار التلاحقات مؤسس على الفكرة الآتية . إذا كان بالعينة عدد قليل جداً من التلاحقات ، مثلا :

ك ك ك ك ك ك ك ك   ث ث ث ث ث ث

فإننا نشك في وجود تجمعات معينة أو نمط معين pattern في عملية الاختيار إذ تشير هذه الحالة إلى أن عملية اختيار العينة لم تكن عشوائية بل تبدأ بالذكور ثم بالإناث .

كذلك إذا كان هناك عدد كبير جداً من التلاحقات ، مثلا

ك ث ك ث ك ث ك ث ك ث ك ث ك ث

فإننا نشك في وجود نوع من النمط الذى يتكرر دورياً وتشير هذه الحالة أيضاً إلى أن الاختيار لم يكن عشوائياً .

ولذلك يبني اختبار التلاحقات على المتغير العشوائي س الذى يعبر عن عدد التلاحقات في المتتابعة التي نتجت في عملية الاختيار . ولهذا المتغير توزيع معروف ، وسطه الحسابي وتباينه كالآتي :

$$\mu_{س} = \frac{٢ ن_١ ن_٢}{ن_١ + ن_٢} + ١ \qquad (١)$$

$$\sigma^{٢}_{س} = \frac{٢ ن_١ ن_٢ (٢ ن_١ ن_٢ - ن_١ - ن_٢)}{(ن_١ + ن_٢)^{٢} (ن_١ + ن_٢ - ١)} \qquad (٢)$$

ويمكن إثبات أنه إذا كانت كل من $ن_١$ ، $ن_٢$ لا تقل عن ١٠ فإن الإحصاءة

$$ص = \frac{س - \mu}{\sigma_{س}} \qquad (٣)$$

يكون توزيعها قريباً من التوزيع المعتدل المعيارى .

ونظرا لأن البيانات التي نحصل عليها تكون بيانات عن متغير وثاب س بينما

التوزيع المعتدل هو توزيع لمتغير متصل فإن الصيغة (٣) تصحح كالآتي تعويضاً عن هذا الاختلاف .

$$ص = \frac{(س \pm \frac{1}{2}) - \mu_{س}}{\sigma_{س}} \qquad (٤)$$

وتؤخذ العلامة السالبة إذا كان عدد التلاحقات (س) في العينة يزيد عن الوسط الحسابي $\mu_{س}$ المحسوب من الصيغة (١) ، وتؤخذ العلامة الموجبة إذا كانت س تقل عن $\mu_{س}$. هذا ويمكن الاستغناء عن هذا التصحيح وإهمال النصف إذا كانت العينة كبيرة .

وحين تكون كل من $ن_{١}$ ، $ن_{٢}$ أكبر أو تساوى عشرة نكون أمام واحد من الحالات الثلاثة الآتية :

(أ) إذا كان الفرض الصفرى هو أن العينة عشوائية والفرض الآخر هو العكس فإننا نستخدم اختباراً ذا جانبين ونرفض الفرض الصفرى عند مستوى الدلالة ٠,٠٥ إذا وقعت القيمة المشاهدة للإحصاءة (٤) خارج المنطقة :

(١,٩٦ > ص > − ١,٩٦) (٥)

ونرفضه عند المستوى ٠,٠١ إذا وقعت خارج المنطقة :

(٢,٥٨ > ص > − ٢,٥٨) (٦)

وهذا بحسب ما جاء بالمثال (٤ − ٣) في الفصل الرابع . ويمكن أن نوجد المناطق المناظرة لأى مستوى دلالة آخر من جدول المساحات أسفل المنحني المعتدل المعيارى .

(ب) أما إذا كان عدد التلاحقات صغيراً وأردنا اختبار الفرض عن وجود نمط يفسر قلة هذه التلاحقات فإن الاختبار في هذه الحالة يكون ذا جانب واحد هو الجانب الأيسر ونرفض الفرض الصفرى عن عشوائية العينة لصالح هذا الفرض عند مستوى الدلالة ٠,٠٥ إذا كانت القيمة المشاهدة للإحصاءة (ع) أقل من — ١,٦٤ ونرفضه عند مستوى الدلالة ٠,٠١ إذا كانت تقل عن – ٢,٣٣ . راجع المثال (٤ – ٣) .

(جـ) كذلك إذا كان عدد التلاحقات كبيراً وأردنا اختبار الفرض عن وجود نمط دورى فإن الاختبار يكون أيضاً ذا جانب واحد هو الجانب الأيمن ونرفض الفرض الصفرى عن عشوائية العينة لصالح هذا الفرض عند المستوى ٠,٠٥ (أو ٠,٠١) إذا كانت القيمة المشاهدة للإحصاءة (ع) تزيد عن ١,٦٤ (أو ٢,٣٣) .

مثال (١٤ – ١) :

أخذت عينة من ٣٠ شجرة دَرْداء كانت قد زرعت من عدة سنوات على طريق زراعى فوجدت المتتابعة الآتية ، حيث ص تعبر عن أن الشجرة مصابة بمرض معين ، ح تعبر عن أنها غير مصابة . اختبر عند مستوى الدلالة ٠,٠١ الفرض بوجود تجمعات أى أن الأشجار المصابة تتجمع معاً .

ح ح ح ح ح   ص ص ص   ح ح ح ح ح ح ح ح ح

ص ص   ح ح   ص ص ص ص ص   ح ح ح ح

الحل :

لدينا س = ٧ (عدد التلاحقات) ، $ن_١$ = ٢٠ (عدد الأشجار السليمة) ،
$ن_٢$ = ١٠ (عدد الأشجار المصابة) .

نظراً لأن كلا من $ن_١$ ، $ن_٢$ لا تقل عن عشرة ، يمكن اعتبار توزيع الإحصاءة (ع) معتدلا معيارياً .

من (١) : $\mu_{\text{س}} = \frac{٢ \times ٢٠ \times ١٠}{٢٠ + ١٠} + ١ = ١٤,٣٣$

من (٢) : $\sigma_{\text{س}} = \sqrt{\frac{٢ \times ٢٠ \times ١٠ \,(٢ \times ٢٠ \times ١٠ - ٢٠ - ١٠)}{٣٠^{٢} \times (٢٠ + ١٠ - ١)}} = ٢,٣٨$

بالتعويض في (٤) : $\text{ص} = \frac{(٧ + \frac{١}{٢}) - ١٤,٣٣}{٢,٣٨} = -٢,٨٧$

الفرض الصفرى : العينة عشوائية .

الفرض الآخر : يوجد تجمعات . (وإذن الاختبار ذو جانب واحد هو الجانب الأيسر)

بما أن − ٢,٨٧ < − ٢,٣٣ نرفض الفرض الصفرى عند مستوى الدلالة ٠,٠١ عن عشوائية العينة ويكون لدينا دليل قوى على أن الأشجار المصابة تقع في تجمعات غير عشوائية .

حالة البيانات الكمية .

لاختبار عشوائية العينة في حالة البيانات الكمية كأن يكون لدينا متتابعة تعبر عن أوزان حيوان ينمو سجلت فى فترات يومية أو أسبوعية ، نستخدم نفس الأسلوب السابق إلا أننا في هذه الحالة نقسم ما لدينا من قيم إلى صنفين بحسب وقوعها **فوق الوسيط أو تحت الوسيط** فنضع حرفاً ا مثلا (أو علامة +) لكل قيمة تزيد عن الوسيط وحرفاً ب (أو العلامة −) لكل قيمة تقل عن الوسيط مع الاحتفاظ بترتيب هذه القيم . وإذا وجدت قيم تساوى الوسيط فإنها تهمل وكأنها لم تكن .

(نذكر أنه لإيجاد الوسيط لمجموعة من الأعداد نرتب هذه الأعداد تصاعدياً أو تنازلياً ثم نأخذ العدد الذى في الوسط إذا كان عدد هذه الأعداد فردياً ، أو متوسط العددين الأوسطين إذا كان عدد الأعداد زوجياً) .

إن طريقة التلاحقات أعلى وأسفل الوسيط تفيد على وجه الخصوص في حالتين رئيسيتين أولهما اختبار الاتجاهات وثانيهما اختبار الأنماط الدورية . فإذا بدأت متتابعة التلاحقات بحروف أغلبها ا ثم بحروف أغلبها ب فإن هناك اتجاهاً إلى أسفل ، وإذا بدأت بحروف أغلبها ب ثم بحروف أغلبها ا فإن هناك ميلا إلى أعلى . أما إذا كان الحرفان ا ، ب يتبادلان بشكل منتظم فإن هذا يشير إلى وجود نمط دورى .

مثال (١٤ – ٢) :

أخذ قطاع على أرض متملحة ، وعند نقط محددة منه قدرت نسب الغطاء النباتي لنوع من النبات بعرض ٥ سم من القطاع . سجلت ٤٠ من هذه النسب بالترتيب كما يلى :

٢٤	٢٧	٢٧	٣٣	٢٨	٢١	١٨	٣٠	٣٥	٢٦	٢٧
٢٤	٢٩	٣٢	٢٣	١٩	٣١	٢٨	٢٨	٢٦	٢٤	١٩
١٤	٢٢	٢٣	٢٢	٢٣	٢٨	١٧	٢١	٢٣	١٣	٢٤
٢٨	٢٠	١٨	٢١	٢١	١٠	١٧				

اختبر ما إذا كان هناك نمط تجمع للغطاء النباتي .

الحل :

الوسيط في هذه العينة = $\frac{1}{2}$(٢٣ + ٢٤) = ٢٣,٥

(يلاحظ أن إيجاد الوسيط يستلزم أولا ترتيب الأعداد المعطاة تصاعديا أو

تنازليا.) بوضع الرمز ا بدلا من أى عدد يزيد عن ٢٣,٥ والرمز ب بدلا من أى عدد يقل عن ٢٣,٥ نحصل على المتتابعة الآتية :

أ أ أ أ أ ب ب أ أ أ أ أ أ أ ب ب أ أ أ أ أ ب ب ب ب ب ب ب ب
أ ب ب ب ب أ أ ب ب ب ب ب ب

يبدو أن القيم الأكبر من الوسيط تميل إلى التجمع في أحد جانبي المتتابعة كما تميل القيم الأصغر من الوسيط إلى التجمع في الجانب الآخر غير أن الحكم الموضوعى على ذلك يبني على اختبار التلاحقات كما يلى :

عدد التلاحقات س = ١٠ ، $ن_١$ = ٢٠ ، $ن_٢$ = ٢٠

$$\text{من (١)}\quad \mu_{س} = \frac{٢ \times ٢٠ \times ٢٠}{٢٠ + ٢٠} + ١ = ٢١$$

$$\text{من (٢)}\quad \sigma^{٢}_{س} = \frac{٢ \times ٢٠ \times ٢٠\ (٢ \times ٢٠ \times ٢٠ - ٢٠ - ٢٠)}{٤٠ \times ٤٠\ (٤٠ - ١)} = ٩,٧٤٣٦$$

$$\therefore \quad \sigma_{س} = ٣,١٢$$

$$\text{من (٤) :}\quad ص = \frac{١٠ + ٠,٥ - ٢١}{٣,١٢} = -٣,٣٦٥$$

الفرض الصفرى : العينة عشوائية (لا يوجد أى نمط)
الفرض الآخر : هناك نمط تجمع ، وإذن الاختبار ذو جانب واحد .
بما أن − ٣,٣٦٥ أصغر من − ٢,٣٣ نرفض الفرض الصفرى عند مستوى الدلالة ٠,٠١ ونحكم بأن هناك نمط تجمع للغطاء النباتي .

حالة العينات الصغيرة :

إذا كان أحد العددين $ن_1$ ، $ن_2$ أو كلاهما أصغر من ١٠ فلا يجوز التقريب بالتوزيع المعتدل المعياري . وفي هذه الحالة نستخدم جداول خاصة كالجدول (١٣) بملحق هذا الكتاب الذى يعطى احتمال أن تقل عدد التلاحقات عن عدد معين $س$ (على أساس صحة الفرض الصفرى عن عشوائية العينة) عند زوج مرتب من الأعداد ($ن_1$ ، $ن_2$) أى يعطى الاحتمال :

ل ($س \geqq س_0$ | ف صحيح) عند ($ن_1$ ، $ن_2$) وحيث $ن_1 \geq ن_2$.

وقد أعد هذا الجدول بحيث يكون الإحداثي الأول $ن_1$ في الزوج المرتب ($ن_1$ ، $ن_2$) أصغر من الإحداثي الثاني $ن_2$ ، أى أننا نرمز بالرمز $ن_1$ لعدد مرات ظهور الحرف الذى يتكرر أقل سواء كان هذا الحرف هو ا أو ب . أما إذا كان $ن_1 = ن_2$ فلا يوجد أى شرط . ونرفض الفرض الصفرى عن عشوائية العينة عند المستوى α (٠,٠٥ مثلا) إذا كان الاحتمال الناتج يقل عن α وإلا نقبل الفرض الصفرى .

مثال (١٤ - ٣) :

ضبطت آلة لكى تصب مقداراً معيناً من سائل ما في كل وعاء يمر تحتها . وجد أنه في ١٥ وعاء متتالياً كانت مقادير السائل باللترات كالآتي :

٣,٦ ٣,٩ ٤,١ ٣,٦ ٣,٨ ٣,٧ ٣,٤ ٤,٠

٣,٨ ٤,١ ٣,٩ ٤,٠ ٣,٨ ٤,٢ ٤,١

هل نستطيع القول بأن المقادير التي توزعها الآلة تتغير عشوائياً ؟

الحل :

الوسيط = ٣,٩

بوضع ا بدلا من كل عدد يزيد عن ٣,٩ ، ب بدلا من كل عدد يقل عن ٣,٩ وإهمال العددين المساويين للعدد ٣,٩ نحصل على المتتابعة الآتية :

ب أ ب ب ب أ ب أ أ ب أ أ

لدينا س = ٨ (عدد التلاحقات) ، $ن_١$ = ٦ ، $ن_٢$ = ٧

مع ملاحظة أن $ن_١$ أصغر من $ن_٢$ وأن كلا منهما أصغر من ١٠

من الجدول (١٣) عند ($ن_١$ ، $ن_٢$) = (٦ ، ٧) ، س = ٨ نجد أن :

ل (س ⩾ ٨ | ف صحيح) = ٠,٧٣٣

وهذا الاحتمال يزيد عن ٠,٠٥ ولذلك نقبل الفرض الصفرى أن حدود المتتابعة تتغير عشوائياً .

(١٤ – ١ – ١) تطبيق آخر لاختبار التلاحقات :

يستخدم اختبار التلاحقات في الكشف عن دلالة تأثير معالجة ما على متغير ما كما يتبين من المثال الآتي .

مثال (١٤ – ٤) :

لمعرفة تأثير هورمون ما على أطوال براعم أحد النباتات أخذت عينة عشوائية من ٢٢ من هذا النبات وقسمت عشوائيا إلى قسمين بكل منهما ١١ نباتا ووضعت النباتات تحت نفس الظروف فيما عدا أن نباتات أحد القسمين عولجت بالهورمون وتركت نباتات القسم الآخر دون معالجة (مجموعة مراقبة) . وبعد أسبوعين وجد أن أطوال البراعم بالملليمترات كالآتي :

المجموعة المعالجة : ٢٦ ٢٨ ٣٧ ٤١ ٥٧ ٥٨ ٦٧ ٧٨ ٨٠ ١٢١ ١٢٢

مجموعة المراقبة : ١٥ ٣٠ ٣٠ ٦٥ ٩١ ١٠٣ ١٣٢ ١٣٥ ١٣٧ ١٤١ ١٩٨

ابحث ما إذا كان الهورمون يعيق نمو براعم هذا النبات .

الحل :

نضم جميع القيم المشاهدة في المجموعتين في متتابعة واحدة ونكتب عناصر هذه المتتابعة بعد ترتيبها ترتيبا تصاعديا ثم نضع الحرف ا تحت كل عنصر من عناصر المجموعة المعالجة والحرف ب تحت كل عنصر من عناصر مجموعة المراقبة كما يلي :

١٥	٢٦	٢٨	٣٠	٣٠	٣٧	٤١	٥٧	٥٨	٦٥	٦٧	٧٨
ب	أ	أ	ب	ب	أ	أ	أ	أ	ب	أ	أ

٨٠	٩١	١٠٣	١٢١	١٢٢	١٣٢	١٣٥	١٣٧	١٤١	١٩٨
أ	ب	ب	أ	أ	ب	ب	ب	ب	ب

الفرض الصفرى : المعالجة ليس لها تأثير في نمو البراعم . أى أن البيانات مأخوذة من مجتمع واحد وبالتالى تتتابع الرموز ا ، ب عشوائيا .

الفرض الآخر : المعالجة تعيق نمو البراعم ، وإذن الاختبار ذو جانب واحد .

إذا كان الفرض الصفرى صحيحا فإن جميع المتتابعات من الرمزين ا ، ب التي يمكن أن تسفر عنها التجربة تكون متساوية الاحتمال ويتوزع هذان الرمزان عشوائيا . أما إذا كانت المجموعتان هما عينتان من مجتمعين مختلفين نتيجة لتأثير المعالجة بالهورمون فإن عناصر كل من المجموعتين تميل إلى التجمع معا ويكون عدد التلاحقات قليلا . وعلى ذلك يمكن استخدام اختبار التلاحقات بالأسلوب السابق بيانه في الأمثلة الثلاثة السابقة .

عدد التلاحقات س = ٩ ، $ن_١$ = ١١ ، $ن_٢$ = ١١

من (١) : $\mu_ر = \frac{٢ \times ١١ \times ١١}{٢٢} + ١ = ١٢$

من (٢) : $\sigma^{2}_{س}$ = $\frac{٢ \times ١١ \times ١١ (٢٤٢ - ٢٢)}{٢٢ \times ٢٢ \times ٢١}$ = ٥,٢٣٨

ومنها $\sigma_{س}$ = ٢,٢٨٨٧

من (٤) : ص = $\frac{٩ + ٠,٥ - ١٢}{٢,٢٨٨٧}$ =– ٠,٨٦٦

بما أن – ٠,٨٦٦ أكبر من – ١,٦٤ لا يكون لدينا دليل ضد الفرض الصفرى عند مستوى الدلالة ٠,٠٥ وإذن لا نستطيع القول من واقع هذه التجربة أن الهرمون يعيق نمو براعم النبات .

تمارين (١٤ – ١)

(١) المتتابعة الآتية تعبر عن الوحدات المعيبة أ والوحدات غير المعيبة ب التي صنعتها آلة ما بالترتيب :

ب ب أ أ أ ب ب ب أ أ ب ب ب ب ب ب ب ب أ أ أ ب ب ب أ
ب ب ب ب أ أ أ

اختبر عند مستوى الدلالة ٠,٠٥ ما إذا كان بالإمكان النظر إلى هذه البيانات على أنها عينة عشوائية .

(٢) الأعداد الآتية هى أعداد الطلاب الذين تغيبوا عن مدرسة في ٢٤ يوماً متتالياً

٣٣ ٣١ ٢٩ ٣٥ ٣١ ٣٣ ٢٨ ٣٠ ٢٨ ٣١ ٢٥ ٢٩
٢٧ ٣٨ ٣١ ٣٢ ٢٨ ٣٠ ٢٦ ٣٣ ٣٠ ٣٦ ٢٨ ٣٥

اختبر العشوائية عند مستوى الدلالة ٠,٠١

(٣) اختبر عند مستوى الدلالة ٠,٠٥ ما إذا كانت قيم العينة الآتية مرتبة ترتيباً عشوائياً :

٩٧	٣	١٦	١٢	٥٥	١٦	٨٤	٦٣	٣٣	٥٧
١٨	٢٦	٢٣	٥٢	٣٧	٧٠	٥٦	٩٩	١٦	٣١

(١٤ - ٢) اختبار الإشارة : SIGN TEST

حين نستخدم اختبار ت لاختبار فرض عن الوسط الحسابي لمجتمع (البند ٦ - ٦ - ١) وعند اختبار فرض تساوى متوسطى مجتمعين (البند ٦ - ٦ - ٣) نشترط أن تكون المجتمعات معتدلة . أما إذا كان هذا الشرط غير متحقق ولا يمكن الدفاع عنه فلا مفر من الالتجاء إلى الاختبارات غير البارامترية . ولعل أسهل وأسرع اختبار لذلك هو الاختبار المعروف باختبار الإشارة ، وهو يبنى على توزيع ذى الحدين .

(١٤ - ٢ - ١) اختبار فرض عن متوسط مجتمع :

نفرض أننا حصلنا على عينة عشوائية من مجتمع متصل ونريد أن نختبر ما إذا كان لهذا المجتمع وسط حسابي معين μ = ا . يبدأ اختبار الإشارة بوضع الإشارة + بدلا من كل قيمة في العينة تزيد عن ا ووضع الاشارة - بدلا من كل قيمة تقل عن ا وإهمال القيم التي تساوى ا . إذا كان الفرض الصفرى μ = ا صحيحاً وكان المجتمع متماثلا نتوقع أن يكون عدد الإشارات الموجبة مساوياً على وجه التقريب لعدد الإشارات السالبة . أما إذا بدا أن أحدهما أكبر مما ينبغى فإننا نرفض ذلك الفرض الصفرى .

ليكن $ن_+$ ، $ن_-$ رمزين لعددى الإشارات الموجبة والسالبة على الترتيب . إذا كان الفرض صحيحاً فإن احتمال أن تزيد أى قيمة مشاهدة عن العدد ا يساوى

احتمال أن تقل عن l وعلى ذلك فإن كلا الاحتمالين يساوى $\frac{1}{2}$ ولذلك فإن اختبار الإشارة يعرف متغيراً عشوائياً س يعبر عن عدد الإشارات الموجبة (أو السالبة) في ن من العناصر. وإذا كانت القياسات مستقلة يكون لهذا المتغير توزيع ذى الحدين دليلاه ن ، $\frac{1}{2}$ حيث ن هو حجم العينة بعد استبعاد القيم التي أهملت. ولوضع قاعدة روتينية لهذا الاختبار نرمز إلى قيم هذا المتغير بالرمز سَ حيث سَ أصغر العددين $ن_+$ ، $ن_-$ أى أن :

$$سَ = أصغر \; (ن_+ ، ن_-)$$

ثم نحسب احتمال أن يأخذ المتغير س قيماً تساوى أو تقل عن القيمة المشاهدة سَ أى نحسب الاحتمال :

$$ل \; (س \leqslant سَ) \qquad (٧)$$

على أساس صحة الفرض الصفرى أن ح = $\frac{1}{2}$ (وبالتالى $\mu = l$) . وإذا كانت α هى مستوى الدلالة الذى اخترناه فإننا نرفض الفرض الصفرى عند هذا المستوى في حالة الاختبار ذى الجانب الواحد إذا كان :

$$ل \; (س \leqslant سَ) < \alpha \qquad (٨)$$

ونقبله إذا زاد هذا الاحتمال عن α أو كان مساوياً لها . وفي حالة الاختبار ذى الجانبين نرفض الفرض الصفرى إذا كان :

$$٢ \; ل \; (س \leqslant سَ) < \alpha \qquad (٩)$$

وحين يكون حجم العينة صغيراً (ن $\leqslant$ ١٥) نوجد الاحتمال (٧) مباشرة من أحد جداول احتمالات توزيع ذى الحدين كالجدول (٣) بملحق هذا الكتاب ، أما إذا كانت ن أكبر من ١٥ ولا يتسع لها هذا الجدول فإننا نستخدم تقريب التوزيع المعتدل لتوزيع ذى الحدين الذي مر بنا بالبند (٤ – ٦) إذا توفرت شروطه ، ونستخدم نفس مناطق الرفض كما في البند (١٤ – ١) الأخير .

مثال (١٤ – ٥) :

الأعداد الآتية هى ١٥ قياساً لمعدلات الأوكتين في نوع من الجاسولين :

٩٢,٧ ٩٤,٨ ٩٦,٩ ٩٩,٨ ٩٦,٣ ٩٥,٥ ٩٦,٨ ٩٤,٧ ٩٧
٩٤,٤ ٩٨,٠ ٩٧,٧ ٩٧,١ ٩٥,٦ ٨٦,٦

اختبر عند مستوى الدلالة ٠,٠١ الفرض الصفرى أن متوسط معدل الأوكتين هو μ = ٩٨ ضد الفرض الآخر أن μ < ٩٨ .

الحـــل :

باستخدام قاعدة الإشارات سابقة الذكر نحصل على الإشارات الآتية وعددها ١٤ إشارة بعد إهمال العدد الذى يساوى ٩٨ .

– – – – – + – – – – + – – – – –

إذن $ن_+$ = ٢ ، $ن_-$ = ١٢ من ١٤ إشارة .
نأخذ $س_{ن}$ = ٢ (أصغر العددين ٢ ، ١٢) .

نختبر الفرض أن توزيع عدد الإشارات الموجبة هو توزيع ذى الحدين دليلاه ١٤ ، ٠,٥ . إذا كان هذا الفرض صحيحاً نجد من الجدول (٣) عند ن = ١٤ ، ح = $\frac{١}{٢}$ أن :

ل (س ≥ ٢) = ٠,٠٠١ + ٠,٠٠٦ = ٠,٠٠٧

وهذا الاحتمال أصغر من مستوى الدلالة ٠,٠١ ولهذا نرفض الصفرى ونستنتج أن متوسط معدل الأوكتين يقل عن ٩٨ .

ملاحظة (١) :

استخدام تقريب التوزيع المعتدل :

في هذا المثال نظراً لأن ن ح = ن (١ − ح) = ١٤ × $\frac{1}{2}$ = ٧ وهذا العدد أكبر من خمسة ، وحسب الإحصاءة (٦) بالبند (٦ − ٤) يمكن تقريب توزيع ذى الحدين الذى لدينا بالتوزيع المعتدل عن طريق الإحصاءة :

$$ص = \frac{(س \pm \frac{1}{2}) - \mu_س}{\sigma_س}$$

حيث $\mu_س$ = ن ح = ٧

، $\sigma_س = \sqrt{ن\ ح\ (١ - ح)} = \sqrt{١٤ \times \frac{1}{2} \times \frac{1}{2}} = ١,٨٧١$

وعلى أساس صحة الفرض نجد أن :

$$ص = \frac{(٢ + \frac{1}{2}) - ٧}{١,٨٧١} = -٢,٤٠٦$$

وهذا العدد يقل عن − ٢,٣٣ وإذن نرفض الفرض الصفرى عند المستوى ٠,٠١

ملاحظة (٢) :

الواقع أن الذى يختبره اختبار الإشارة هو الوسيط ولكن نظرا لأننا افترضنا أن المجتمع متماثل فإن الوسيط يكون هو نفسه الوسط الحسابى ، أما إذا لم يكن المجتمع متماثلا فإن اختبار الإشارة يكون اختباراً عن **الوسيط** وليس عن الوسط الحسابي .

(١٤ − ٢ − ٢) مقارنة متوسطى مجتمعين غير معتدلين :

نفرض أن لدينا مجموعة من أزواج المشاهدات ($س_ر$ ، $ص_ر$) من مجتمعين

متصلين غير معتدلين ونريد اختبار ما إذا كان لهذين المجتمعين وسطان حسابيان متساويان ($\mu_1 = \mu_2$). بنفس أسلوب البند السابق ، يبدأ اختبار الإشارة بوضع الإشارة + أو الإشارة − لكل فرق $ف_ر = س_ر - ص_ر$ بحسب كون هذا الفرق موجباً أو سالباً ، وإهمال الحالات التي ينعدم فيها هذا الفرق .

إذا كان الفرض الصفرى $\mu_1 = \mu_2$ صحيحاً وكان كل من المجتمعين متماثلا ينبغى أن يكون مجموع الإشارات الموجبة في العينة مساويا بالتقريب لمجموع الإشارات السالبة ويكون احتمال أن يكون الفرق ف موجباً يساوى احتمال أن يكون هذا الفرق سالباً ويكون كلا الاحتمالين مساوياً للعدد $ح = \frac{1}{2}$ ولذلك فإن الاختبار يعرف متغيراً عشوائياً له توزيع ذى الحدين دليله $ح = \frac{1}{2}$ بشرط صحة الفرض الصفرى . ويسير الاختبار بنفس الأسلوب المذكور في البند السابق

مثال (١٤ − ٦) .

في دراسة لمعرفة تأثير نظام جديد في المرور جمعت البيانات الآتية عن عدد الحوادث التي وقعت في ١٢ تقاطعاً من التقاطعات الخطره (على فرض أنها عينة عشوائية) خلال ٣ شهور قبل النظام الجديد و٣ شهور بعده .

قبل النظام الجديد : ٣ ٥ ٢ ٣ ٣ ٦ ٠ ٤ ١ ٥ ٤

بعد النظام الجديد : ٠ ١ ٢ ٢ ١ ١ ٢ ٣ ٣ ٣ ٠

ابحث عند مستوى الدلالة ٠,٠٥ ما إذا كان النظام الجديد له اثر في تقليل الحوادث عند تلك التقاطعات

الحل :

باستخدام قاعدة الإشارات للفروق نحصل على الإشارات الآتية وعددها ١٢

$$+ \; + \; + \; + \; + \; + \; - \; + \; - \; - \; + \; -$$

$ن_+ = ١٠$ ، $ن_- = ٢$ من ١٢ إشارة

الفرض الصفرى ف. : $\mu_1 = \mu_2$ متوسط عدد الحوادث واحد في الحالتين

ف$_1$: $\mu_1 > \mu_2$

نأخذ س. = ٢ (أصغر العددين ١٠ ، ٢)

نختبر الفرض أن توزيع عدد الإشارات السالبة هو توزيع ذى الحدين دليلاه ١٢ ، $\frac{1}{2}$. إذا كان هذا الفرض صحيحاً نجد من الجدول (٢) عند ن = ١٢ ، ح = $\frac{1}{2}$ أن :

ل (س $\geq$ ٢) = ٠,٠٠٣ + ٠,٠١٦ = ٠,٠١٩

وهذا الاحتمال أقل من ٠,٠٥ وإذن نرفض الفرض الصفرى عند مستوى الدلالة ٠,٠٥ لصالح الفرض الآخر ونحكم بأن النظام الجديد قد أثر في تقليل عدد الحوادث عند التقاطعات الخطرة .

تمارين (١٤ – ٢)

(١) في إحدى التجارب المعملية نتجت ١٨ قيمة لمعامل الاحتكاك بين الجلد وأحد المعادن وكانت هذه القيم كالآتي :

٠,٥٩	٠,٤٨	٠,٥٣	٠,٥٧	٠,٥٥	٠,٦١	٠,٥٦	٠,٦٠	٠,٥٥
٠,٥٨	٠,٥٢	٠,٥٥	٠,٥٦	٠,٥١	٠,٥٧	٠,٥٦	٠,٥٩	٠,٥٨

استخدام اختبار الإشارة عند مستوى الدلالة ٠,٠٥ لاختبار الفرض الصفرى أن متوسط معامل الاحتكاك μ = ٠,٥٥ ضد الفرض الآخر أن $\mu \neq$ ٠,٥٥

(٢) الآتي هي أعداد الحفريات التي وجدها اثنان من علماء الآثار في بقايا مساكن أثرية على سطح جبل في ٣٠ يوماً :

العالم الأول :	١	٠	٢	٣	١	٠	٢	٢	٣	٠	١	١	٤	١	٢
العالم الثاني :	٠	٠	١	٠	٢	٠	٠	١	١	٢	٠	١	٢	١	١

العالم الأول : ١ ٣ ٥ ٢ ١ ٣ ٢ ٤ ١ ٣ ٢ ٠ ٢ ٤ ٢
العالم الثانى : ٠ ٢ ٢ ٦ ٠ ٢ ٣ ٠ ٢ ١ ٠ ١ ٠ ١ ٠

استخدم اختبار الإشارة عند مستوى الدلالة ٠,٠١ لاختبار الفرض الصفرى أن العالمين على نفس الكفاءة في العثور على الحفريات ضد الفرض الآخر أن العالم الأول أفضل .

(٣) أعطى كل من ١٠ مرضي نوعان من المهدئات ا ، ب والجدول الآتي يعرض الزيادة في مدة النوم بالساعات . هل الفرق بين نوعى المهدئات ذو دلالة ؟

أ : ١,٩ ٠,٨ ١,١ ٠,١ −٠,١ ٤,٤ ٥,٥ ١,٦ ٤,٦ ٣,٤
ب : ٠,٧ −١,٦ −٠,٢ −١,٢ −٠,١ ٣,٤ ٣,٧ ٠,٨ ٠,٠ ٢,٠

(١٤ – ٣) اختبار ويلكوكسن للمقارنات التزاوجية :
WILCOXON TEST FOR PAIRED COMPARISONS

إن اختبار الإشارة الذى ورد بالبند (١٤ – ٢ – ٢) يحدد أياً من المجتمعين المأخوذة منهما العينات هو الأكبر في المتوسط ولكنه لا يحدد مقدار الفرق بينهما . والاختبار الذى يحدد كلا الاتجاه والمقدار هو ذلك المعروف باختبار ويلكوكسن للمقارنات التزاوجية ، وهو فضلا عن هذا أكثر حساسية من اختبار الإشارة في الكشف عن وجود فرق بين متوسطى مجتمعين توزيعاهما مجهولان . وتتضح أهمية هذا الاختبار في الحالات التي لا تنطبق فيها شروط الاختبار المقدم في البند (٨ – ٧) .

نفرض أننا حصلنا على ن من أزواج القيم (سر ، صر) . وليكن فر = سر – صر هى الفروق بين هذه الأزواج . لاختبار الفرض الصفرى $\mu_1 = \mu_2$ عن تساوى متوسطى المجتمعين يبدأ اختبار ويلكوكسن بإهمال الفروق المساوية للصفر

ثم ترتيب الفروق الباقية بصرف النظر عن إشاراتها أى بحسب القيم المطلقة لهذه الفروق ، فتعطى الرتبة ١ لأصغر فرق في القيمة المطلقة وتعطى الرتبة ٢ للفرق التالى في الصغر وهكذا . وحين تتساوى القيم المطلقة لاثنين أو أكثر من هذه الفروق يعطى لكل منهما متوسط الرتب التي كانت ستعطى لو كانت هذه الفروق متميزة .

إذا كان الفرض الصفرى $\mu_1 = \mu_2$ صحيحا نتوقع أن يكون مجموع الرتب المناظرة للفروق الموجبة في العينة مساويا بالتقريب لمجموع الرتب المناظرة للفروق السالبة . لنرمز إلى هذين المجموعين بالرمزين $م_+$ ، $م_-$ على الترتيب ، وليكن

$$س = \text{أصغر } (م_+ ، م_-)$$

نظراً لأن س تتغير من عينة إلى أخرى فإننا ننظر إليها على أنها قيمة مشاهدة من متغير عشوائي س . إن هذا المتغير له توزيع معروف متوسطه وتباينه كالآتي :

$$\mu_س = \frac{ن}{٤} (ن + ١) \qquad (١٠)$$

$$، \sigma^٢_س = \frac{ن}{٢٤} (ن + ١)(٢ن + ١) \qquad (١١)$$

وعلى ذلك تبني قاعدة الاختبار على إيجاد احتمال الحصول بالصدفة وحدها على قيمة تساوى أو تقل عن قيمة س المشاهدة أى على إيجاد الاحتمال ل (س $\leqslant$ س) . ونظراً لصعوبة إيجاد هذا الاحتمال فقد أعدت جداول لإعطاء القيم الحرجة لتوزيع المتغير س ومنها الجدول (١٤) بملحق هذا الكتاب وهو يعطى القيمة الحرجة عند ن = ٥ ، ٦ ، ... ، ٣٠ عند كل من مستويات الدلالة ٠,٠١ ، ٠,٠٢٥ ، ٠,٠٥ للاختبارات ذات الجانب الواحد وعند كل من مستويات الدلالة ٠,٠٢ ، ٠,٠٥ ، ٠,١٠ للاختبارات ذات الجانبين ، وأى قيمة مشاهدة س تقل عن القيمة المذكورة في الجدول تكون ذات دلالة عند المستوى المذكور وتدعو إلى رفض الفرض الصفرى وذلك لأن هذا الجدول يعطى قيم س حيث الاحتمال ل (س $\leqslant$ س) يكون أصغر من مستوى الدلالة المذكور .

وفي الحالة التي تزيد فيها ن عن ٣٠ ولا يتسع لها الجدول (١٤) نستخدم الإحصاءة .

$$ص = \frac{س - \frac{ن}{٤}(ن + ١)}{\sqrt{\frac{ن}{٢٤}(ن + ١)(٢ن + ١)}} \qquad (١٢)$$

التي يقترب توزيعها من التوزيع المعتدل المعيارى .

مثال (١٤ – ٧) :

الأعداد المدونة بالعمودين الثاني والثالث من الجدول الآتي هي متوسطات أعداد المواليد في المرة الواحدة لسلالتين من فيران التجارب كانتا محفوظتين في مستعمرات كبيرة في الولايات المتحدة ، وذلك في الأعوام التسعة من ١٩١٦ إلى ١٩٢٤ :

السنة	السلالة (أ)	السلالة (ب)	ف	الرتب (مع إهمال الإشارة)
١٩١٦	٢,٦٨	٢,٣٦	+ ٠,٣٢	٩
١٩١٧	٢,٦٠	٢,٤١	+ ٠,١٩	٨
١٩١٨	٢,٤٣	٢,٣٩	+ ٠,٠٤	٢
١٩١٩	٢,٩٠	٢,٨٥	+ ٠,٠٥	٣
١٩٢٠	٢,٩٤	٢,٨٢	+ ٠,١٢	٧
١٩٢١	٢,٧٠	٢,٧٣	– ٠,٠٣	١
١٩٢٢	٢,٦٨	٢,٥٨	+ ٠,١٠	٦
١٩٢٣	٢,٩٨	٢,٨٩	+ ٠,٠٩	٥
١٩٢٤	٢,٨٥	٢,٧٨	+ ٠,٠٧	٤

اختبر عند مستوى الدلالة ٠,٠١ ما إذا كان متوسط الخلفة للسلالة أ يزيد عنه في السلالة ب .

الحل :

نحسب الفروق فـ بين كل زوج من المشاهدات (السلالة أ - السلالة ب) مع الاحتفاظ بالإشارة الموجبة أو السالبة كما هو مبين بالعمود الرابع . نرتب الفروق من الأصغر إلى الأكبر بصرف النظر عن الإشارة كما هو مبين بالعمود الأخير .

لدينا ن = ٩ (عدد أزواج القيم في العينة)

م+ = ٤٤ (مجموع الرتب المناظرة للفروق الموجبة)

م- = ١ (مجموع الرتب المناظرة للفروق السالبة)

نأخذ س = ١ (أصغر العددين ٤٤ ، ١)

من الجدول (١٤) وعند ن = ٩ ، α = ٠,٠١ نجد القيمة الحرجة ٣ وأى قيمة س تساوى أو تقل عن هذه القيمة تكون ذات دلالة عند المستوى ٠,٠١ وبما أن س = ١ $\leqslant$ ٣ فإنها تكون ذات دلالة وتدعو إلى رفض الفرض الصفرى ونستنتج أن متوسط الخلفة في السلالة (أ) أكبر منه في السلالة (ب) .

ملاحظة (١) :

في هذا المثال يحق لنا اعتبار أننا بصدد مقارنات تزاوجية وذلك بملاحظة توازى التغيرات في السلالتين معاً خلال السنوات التسع . ففي السنتين ١٩١٧ ، ١٩١٨ وهما سنتا حرب في الولايات المتحدة وفى العالم كله ، أدى النقص في الرعاية وفي الغذاء إلى قلة عدد الذرية في السلالتين ثم تحسن هذا العدد بمجرد تحسن الظروف . كذلك نلاحظ أنه في العام ١٩٢٢ كان هناك هبوط في الذرية في كلا السلالتين ، مما يشير إلى أن التغيرات ترجع إلى أسباب بيئية . ولهذا يكون من المناسب تناول هذه البيانات على أنها مقارنات تزاجية العامل الثابت فيها هو عامل السلالة أما السنوات فهى عامل التكرارات .

ملاحظة (٢) :

يمكن استخدام اختبار ويلكوكسن للمقارنات التزاوجية لاختبار الفرض

الصفرى أن الفرق بين متوسطى مجتمعين يأخذ قيمة معينة ، مثلا $\mu_1 - \mu_2 = $ أ .
ولا يختلف الإجراء المطلوب عن الإجراء السابق إلا في أننا نطرح العدد أ من كل فرق ف قبل إعطاء الرتب كما في المثال الآتى .

مثال (١٤ - ٨) :

كان هناك ادعاء بأنه إذا أعطى طالب في المستوى الرابع عينات من امتحانات سابقة فإن درجاته عند التخرج تزيد بمقدار ٥٠ نقطة . اختير ٢٠ طالباً عشوائياً في هذا المستوى وقسموا إلى ١٠ أزواج بحيث كان الطالبان في كل زوج متساويين تقريباً في معدلهما التراكمى عن السنوات الثلاثة السابقة ، وأعطيت عينات من امتحانات سابقة لواحد فقط من كل زوج (اختير بطريقة عشوائية) قبل الامتحان بأسبوع وسجلت درجات الامتحان النهائي في العمودين الثاني والثالث من الجدول الآتي . اختبر عند مستوى الدلالة ٠,٠٥ الفرض الصفرى أن $\mu_1 - \mu_2 = $ ٥٠ ضد الفرض الآخر أن $\mu_1 - \mu_2 < $ ٥٠

الزوج	بامتحانات سابقة	بغير امتحانات سابقة	ف	ف - ٥٠	الرتب (مع إهمل الاشارة)
١	٥٣١	٥٠٩	٢٢	-٢٨	٥
٢	٦٢١	٥٤٠	٨١	٣١	٦
٣	٦٣٣	٦٨٨	-٢٥	-٧٥	٩
٤	٥٧٩	٥٠٢	٧٧	٢٧	٣,٥
٥	٤٥١	٤٢٤	٢٧	-٢٣	٢
٦	٦٦٠	٦٨٣	-٢٣	-٧٣	٨
٧	٥٩١	٥٦٨	٢٣	-٢٧	٣,٥
٨	٧١٩	٧٤٨	-٢٩	-٧٩	١٠
٩	٥٤٣	٥٣٠	١٣	-٣٧	٧
١٠	٥٧٥	٥٢٤	٥١	١	١

لدينا ن = ١٠ (حجم العينة) .

، $م_+$ = ٦ + ٣,٥ + ١ = ١٠,٥

، $م_-$ = ٥ + ٩ + ٢ + ٨ + ٣,٥ + ١٠ + ٧ = ٤٤,٥

تأخذ س = ١٠,٥ (أصغر العددين ١٠,٥ ، ٤٤,٥)

من الجدول (١٤) وعند ن = ١٠ ، α = ٠,٠٥ نجد القيمة الحرجة ١١.

بما أن س = ١٠,٥ ≤ ١١ نرفض الفرض الصفرى ونستنتج أنه (في المتوسط) إعطاء الامتحانات السابقة لا يجعل درجة التخرج تزيد بمقدار يصل إلى ٥٠ نقطة .

تمارين (١٤ – ٣)

(١) بالجدول الآتي الأوزان بالكيلو جرامات لخمسة أشخاص قبل أن يمتنعوا عن التدخين وبعد ٥ أسابيع من امتناعهم عنه :

	(١)	(٢)	(٣)	(٤)	(٥)
قبل	٦٦	٨٠	٦٩	٥٢	٥٤
بعد	٧١	٨٢	٦٨	٥٦	٥٩

استخدم اختبار ويلكوكسن للمقارنات التزاوجية عند مستوى الدلالة ٠,٠٥ لاختبار أن الامتناع عن التدخين ليس له تأثير في زيادة الوزن ضد الفرض الآخر أنه يزيد الوزن .

(١٤ – ٤) اختبار الاتجاه : TEST FOR TREND

ليكن ص متغيراً عشوائياً ، س متغيراً رياضياً (كما في الفصل التاسع عن الانحدار الخطى البسيط) . كثيراً ما نتساءل : إذا ازدادت قيم س فهل تزداد معها قيم ص (اتجاه موجب) أم تنقص قيم ص (اتجاه سالب) ؟ أى أننا نريد اختبار الفرض الصفرى أنه لا يوجد اتجاه ، ضد الفرض الآخر بوجود اتجاه موجب أو سالب أو بوجود اتجاه بصفة عامة .

إذا توفرت الافتراضات المذكورة في الفصل التاسع (خطية العلاقة بين س ، ص واعتدالية توزيع المتغير العشوائي ص) فإننا نتبع الأسلوب المبين بذلك الفصل ، أما إذا لم تكن متوفرة فيمكننا أن نستخدم اختباراً بسيطاً يعرف باختبار الاتجاه .

وعلى فرض أن لدينا ن من أزواج القيم ($س_ر$ ، $ص_ر$) ناتجة من عينة عشوائية فإن هذا الاختبار يبدأ بترتيب قيم س ترتيباً تصاعدياً ثم ملاحظة قيم ص المناظرة . في حالة وجود اتجاه موجب ينبغي أن تزداد قيم ص مع ازدياد س . أما إذا لم يوجد اتجاه فإن قيم ص تسلك سلوكاً عشوائياً دون أى رتابه أى دون أن يكون هناك نسق معين . والمؤشر لذلك هو عدد الاستبدالات transpositions في المتغير ص أى عدد المرات التي لا تكون فيها قيم ص في مكانها الطبيعى من التزايد أى حين تسبق بعض هذه القيم قيماً أصغر منها . ولذلك فإن الاختبار يعرف متغيراً عشوائياً س يعبر عن عدد الاستبدالات ثم يحسب الاحتمال :

$$ل\,(س \leqslant س_{٠}) \qquad (١٣)$$

على أساس صحة الفرض الصفرى بعدم وجود اتجاه وحيث $س_{٠}$ هى عدد الاستبدالات المشاهدة في العينة . فإذا كان هذا الاحتمال أقل من مستوى الدلالة الذى نختاره فإننا نرفض الفرض الصفرى عند هذا المستوى لصالح الفرض الآخر . وقد أعدت جداول للاحتمالات المتجمعة المبينة بالصيغة (١٣) ومنها الجدول (١٥) بملحق هذا الكتاب الذى يعطى هذه الاحتمالات لقيم ن = ٣ ، ٤ ، ... ، ٢٠ بالنسبة للمتغيرات المتصلة .

مثال (١٤ - ٩) :

في بحث لمعرفة تأثير تعرض بذور الشعير لنوع من الأشعة على المحصول الناتج وجدت البيانات الآتية مع ملاحظة أن س وضعت مرتبة ترتيباً تصاعدياً وأن ص مقاسة بالجرامات في الجوال .

مقدار الأشعــة (س) : ٠ ٢ أ ٤ أ ٤٠ أ ٤٠٠ أ
مقدار المحصول (ص) : ٢٩,٢ ٣٠,٢ ٢٨,٢ ٢٩,٧ ٣٠,٤

اختبر عند مستوى الدلالة ٠,٠٥ ما إذا كان مقدار المحصول يزداد بازدياد مقدار الأشعة .

الحل :

الفرض الصفرى ف٠ : تعريض البذور للأشعة لا يؤثر في المحصول .
الفرض الآخر ف١ : يوجد اتجاه موجب (ص تزداد بازدياد س) .
نحسب عدد الاستبدالات كالآتي :
العدد ٢٩,٢ يسبق العدد ٢٨,٢ (استبدال واحد) .
العدد ٣٠,٢ يسبق العددين ٢٨,٢ ، ٢٩,٧ (استبدالين اثنين) .
إذن عدد الاستبدالات س = ٣ في ن = ٥ من القيم الصادية .
من الجدول (١٥) عند ن = ٥ ، س = ٣ وعلى أساس صحة الفرض الصفرى نجد أن : ل (س ⩾ ٣) = ٠,٢٤٢
وهذا الاحتمال أكبر من مستوى الدلالة ٠,٠٥ وعلى ذلك لا نستطيع رفض الفرض الصفرى ونستنتج أنه في حدود هذه التجربة ليس للأشعة تأثير جوهرى على محصول الشعير .

ملاحظة :

إذا اشتملت العينة على م من القيم المتساوية للمتغير ص نحسب عدد الاستبدالات كما سبق ثم نضيف إليه العدد ١/٤م (م − ١) وهو متوسط الاستبدالات في حالة تباديل م من الأشياء . فمثلا إذا وجدت قيمتان متساويتان أى م = ٢ نضيف العدد ١/٤ × ٢ × ١ = ١/٢ وإذا كانت م = ٣ نضيف العدد ١/٤ × ٣ × ٢ = ٣/٢ لكل ثلاثة قيم متساوية وهكذا .

تمارين (١٤ – ٤)

استخدم اختبار الاتجاه لكل من العينات الآتية :

(١)

س :	١٥	٢٠	٣٠	٤٠	٥٠	٦٠	٧٠
ص :	٦,٥	٥,٦	٥,٤	٦,٠	٤,٦	١,٤	٠,١

حيث ص هي محتوى الأكسجين (ملليجرام / لتر) في أحد البحيرات عند العمق س بالأمتار .

(٢)

س :	١٠	١٥	٣٠	٤٠	٥٠	٥٥	٨٠	١٠٠
ص :	٥٠	٤٦	٤٣	٤٣	٣٦	٣٩	٣٧	٣٣

حيث س هي محتوى الدسلفيد في ميثيل الصوف ،

، ص نسبة تركيز محتوى الماء .

(٣)

س :	١٢١	١٢٠	٩٥	١٢٣	١٤٠	١١٢	٩٢	١٠٠	١٠٢	٩١
ص :	٥٢١	٤٦٥	٣٥٢	٤٥٥	٤٩٠	٣٨٨	٣٠١	٣٩٥	٣٧٥	٤١٨

حيث س تعبر عن ضغط الدم بوحدات معينة ، ص وزن القلب بالجرامات لعينة من ١٠ مرضي ذكور ماتوا بأحد أمراض المخ .

(١٤ – ٥) اختبار كروسكال – واليس KRUSKAL-WALLIS TEST

يستخدم هذا الاختبار كبديل لطريقة تحليل التباين للتجارب ذوات العامل الواحد حين لا تتوفر شروطها .

وتتلخص المشكلة التى نتناولها هنا فيما يلى :

« لدينا ك من العينات العشوائية أحجامها $ن_١$ ، $ن_٢$ ، . . . ، $ن_ك$ مأخوذة

من ك من المجتمعات ، ونرغب فى اختبار الفرض الصفرى ف أن لهذه المجتمعات متوسطات متساوية » . إذا كان هذا الفرض صحيحا يمكن النظر إلى هذه العينات على أنها عينة عشوائية واحدة حجمها ن (هو مجموع أحجام العينات) مأخوذة من مجتمع مشترك . نرتب قيم هذه العينة من الأصغر إلى الأكبر – من ١ إلى ن – ونستعيض عن كل قيمة مشاهدة بالترتيب المناظر لها . نجمع تراتيب وحدات كل عينة على حدة ، ولتكن $\text{ت}_{١}$ ، $\text{ت}_{٢}$ ، ... ، $\text{ت}_{\text{ك}}$ هى مجاميع هذه التراتيب . إذا كان الفرض الصفرى $\text{ف}_{٠}$ صحيحا فإن هذه المجاميع تكون ذات قيم متقاربة ، أما إذا لم يكن $\text{ف}_{٠}$ صحيحا فإن التراتيب الكبرى تميل إلى أن تقع فى العينات المأخوذة من المجتمعات التى لها أكبر المتوسطات . وعلى هذا نكون فى حاجة إلى إحصاءة اختبار تكشف عن وجود أو عدم وجود واحد أو أكثر من المجاميع الترتيبية $\text{ت}_{\text{ى}}$ تكون كبيرة بدرجة لا نتوقع حدوثها عن طريق الصدفة .

وقد وجد أن أحد الاحصاءات التى تصلح لذلك هى تلك التى قدمها كروسكال وواليس وهى تأخذ الصيغة الآتية :

$$\text{ص} = \frac{١٢}{\text{ن}\,(\text{ن} + ١)} \ \text{مج}\ \frac{\text{ت}_{\text{ى}}^{٢}}{\text{ن}_{\text{ى}}} - ٣\,(\text{ن} + ١) \qquad (١٤)$$

حيث ن مجموع حجوم العينات ، $\text{ت}_{\text{ى}}$ مجموع تراتيب وحدات العينة ى ، (ى = ١ ، ٢ ، ... ، ك) . ولهذه الإحصاءة توزيع قريب من توزيع $\chi^{٢}$ بدرجات حرية ك – ١ وبالتالى يمكن استخدام جدول $\chi^{٢}$ لاختبار الفرض الصفرى عن تساوى متوسطات المجتمعات ، فنرفض ف . إذا كانت القيمة المشاهدة للإحصاءة (١٤) أكبر من القيمة الحرجة فى توزيع $\chi^{٢}$ عند مستوى الدلالة الذى نختاره .

مثال (١٤ – ١٠) :

أراد أحد رجال التربية الرياضية اختبار أفضلية ثلاثة طرق جديدة فى تعليم لعبة كرة السلة فأخذ عينة عشوائية من عشرين من المبتدئين فى هذه اللعبة وقسمها عشوائيا إلى أربع مجموعات بكل منها خمسة لاعبين . دربت إحدى المجموعات بالطريقة المعتادة بينما دربت كل من المجموعات الثلاث الأخرى بإحدى الطرق الجديدة . وبعد فترة التدريب قيست مهارات اللاعبين وسجلت درجاتهم فى الجدول الآتى الذى سجلت فيه أيضا التراتيب المناظرة للدرجات (وهى تلك الموضوعة بين الأقواس) . المطلوب اختبار ما إذا كانت هناك فروق ذات دلالة بين مهارات اللاعبين الناتجة عن الطرق الأربع .

	الطرق الجديدة للتدريب			الطريقة المعتادة
	أ	ب	حـ	ء
	(٨) ٦٣,٠	(١٨) ٨٥,٠	(٢٠) ٩٠,٢	(٩) ٦٥,٠
	(٢) ٤٧,٠	(١٦) ٨٠,١	(١١) ٧٠,٧	(١) ٤٥,٢
	(٤) ٥١,٠	(١٥) ٧٩,٠	(١٩) ٨٦,٠	(٣) ٥٠,٩
	(١٣) ٧٤,٠	(١٠) ٦٧,٠	(٧) ٦٢,٣	(١٤) ٧٥,٠
	(٦) ٦٠,٠	(١٧) ٨٢,٣	(١٢) ٧٢,٣	(٥) ٥٨,٨
ت	(٣٣)	(٧٦)	(٦٩)	(٣٢)

الحل

الفرض الصفرى ف : الطرق الأربع تؤدى فى المتوسط إلى نفس الدرجة من المهارات .

من الاحصاءة (١٤) ومع ملاحظة أن ن = ٢٠ ، ك = ٤ نجد أن

$$ص_ر = \frac{١٢}{٢٠ \times ٢١} \left(\frac{٣٣^٢ + ٧٦^٢ + ٦٩^٢ + ٣٢^٢}{٥} \right) - ٣ \times ٢١$$

$$= ٠{,}٠٢٨٥٧ \times ٢٥٣٠ - ٦٣$$

$$= ٩{,}٢٨٢١ \qquad \text{بدرجات حرية ك} - ١ = ٣$$

ولكن $\chi^2_{٠{,}٠٥[٣]} = ٧{,}٨١٥$ ، إذن نرفض ف. عند مستوى الدلالة ٠,٠٥ ، ونحكم بأن المهارات تختلف باختلاف طرق التدريب . ويبدو أن الطريقة ب هى الأفضل .

(١٤ - ٦) اختبار فريدمان FRIEDMAN TEST

يستخدم هذا الاختبار كبديل لطريقة تحليل التباين للتجارب التى تصمم على هيئة قطاعات كاملة التعشية أو للتجارب ذوات العاملين غير المتفاعلين ، حين لا تتوفر شروطها .

نفرض أن لدينا ك من المعالجات ونرغب فى اختبار ما إذا كان لهذه المعالجات تأثيرات مختلفة على وحدات متغير ما ، ونفرض أنه عند تطبيق هذه المعالجات على وحدات التجريب يدخل عامل خارجى له هـ من المستويات قد يؤثر فى النتائج التى نحصل عليها وينبغى إذن استبعاد أثر هذا العامل . لتحقيق هذا الغرض نقوم بتعشية هذا العامل بحيث يكون لكل مستوى من مستويات العامل الذى [illegible] نفس الفرصة للتعرض له . فنسحب ك من الوحدات من كل من الـ هـ مجتمعات التى تمثل مستويات هذا العامل لنحصل على هـ من القطاعات حجم كل منها ك ثم نقوم بتطبيق الـ ك معالجات عشوائيا على وحدات كل قطاع ثم نسجل المشاهدات فى جدول ذى ك من الأعمدة تمثل المعالجات ، هـ من الصفوف تمثل القطاعات

تبدأ طريقة فريدمان بترتيب المشاهدات فى كل قطاع (صف) على حدة من

١ إلى ك بحيث يعطى الترتيب ١ للمشاهدة الأصغر ويعطى الترتيب ك للمشاهدة الأكبر . إذا كان الفرض الصفرى $\text{ف}_{٠}$ صحيحا أى كانت المعالجات لها تأثيرات واحدة على المتغير ، ومع ملاحظة أن المشاهدات فى كل قطاع يمكن اعتبارها عينة عشوائية حجمها ك من مجتمع واحد ، فإن التراتيب العالية تتوزع بين مختلف الأعمدة فى مختلف الصفوف . أما إذا كان $\text{ف}_{٠}$ غير صحيح فإن التراتيب العالية تميل إلى التجمع فى العمود الذى يمثل المعالجة ذات المتوسط الأكبر .

نجمع التراتيب فى كل عمود ولنرمز بالرمز $\text{ت}_{\text{و}}$ لمجموع تراتيب العمود و (و = ١ ، ٢ ، . . . ، ك) . نريد احصاءة اختبار تكشف عن وجود أو عدم وجود واحد أو أكثر من المجاميع الترتيبية تكون كبيرة بدرجة لا تحدث بالصدفة . وقد وجد أن أحدى الاحصاءات الصالحة لهذا الغرض هى تلك التى قدمها فريدمان وهى تأخذ الصيغة الآتية :

$$\text{ص} = \frac{١٢}{\text{هـ ك (ك + ١)}} \sum_{١}^{\text{ك}} \left[\text{ت}_{\text{و}} - \frac{\text{هـ (ك + ١)}}{٢}\right]^{٢} \qquad (١٥)$$

حيث ك عدد الأعمدة ، هـ عدد الصفوف ، $\text{ت}_{\text{و}}$ مجموع التراتيب فى العمود و . وتوزيع هذه الاحصاءة قريب من توزيع $\chi^{٢}$ بدرجات حرية ك − ١ وبالتالى نستطيع استخدام جدول $\chi^{٢}$ فنرفض الفرض الصفرى $\text{ف}_{٠}$ إذا كانت القيمة المشاهدة لهذه الاحصاءة أكبر من القيمة الحرجة لتوزيع $\chi^{٢}$ عند مستوى الدلالة الذى نختاره .

مثال (١٤ − ١١) :

أراد فريق أبحاث المستهلك فى أحد مصانع فرامل الدراجات مقارنة ثلاثة أنواع من الفرامل التى ينتجها . ولما كان نوع الدراجة التى تستخدم فى التجربة قد يؤثر

فى نتائج الدراسة فقد رؤى استبعاد أثر هذا العامل باستخدام تصميم القطاعات كاملة التعشية . اختير ٦ أنواع من الدراجات لتكوين ٦ قطاعات بكل منها ٣ دراجات ووزعت الأنواع الثلاثة من الفرامل عشوائيا على كل قطاع . المتغير الذى تقارن به الفرامل هو الزمن بالأسابيع الذى يمضى حتى يتطلب الأمر اصلاحا أساسيا فيها . وقد سجلت هذه الأزمان فى الجدول الآتى ، كما رتبت هذه الأزمان فى كل قطاع على حدة ووضعت التراتيب المناظرة بين الأقواس فى نفس الجدول .

نوع الدراجة القطاعات	نوع الفرامل ا	ب	حـ
(١)	٥,٢ (٢)	٧,٣ (٣)	٣,٠ (١)
(٢)	٦,٨ (١)	٨,٩ (٣)	٧,٥ (٢)
(٣)	٦,٣ (٢,٥)	٦,٣ (٢,٥)	٦,٠ (١)
(٤)	١٣,٠ (١,٥)	١٤,٨ (٣)	١٣,٠ (١,٥)
(٥)	١٢,٨ (٢,٥)	١٢,٨ (٢,٥)	١١,٠ (١)
(٦)	١٥,٠ (٢)	١٥,٢ (٣)	١٤,٥ (١)
$ت_ى$	(١١,٥)	(١٧)	(٧,٥)

الحل :

الفرض الصفرى $ف_ٍ$: الأنواع الثلاثة من الفرامل ذات عمر واحد .

من (١٥) ، لدينا ك = ٣ ، هـ = ٦ .

$$ص_ى = \frac{١٢}{٦ \times ٣ \times ٤} \sum [ت_ى - \frac{٦ \times ٤}{٢}]^٢$$

$$= \frac{١}{٦} [(١١,٥ - ١٢)^٢ + (١٧ - ١٢)^٢ + (٧,٥ - ١٢)^٢]$$

$$= \frac{١}{٦} \times ٤٥,٥ = ٧,٥٨$$

ولكن $\chi^٢_{٠,٠٥[٢]} = ٥,٩٩١$, إذن نرفض الفرض الصفرى عند مستوى الدلالة ٠,٠٥ ونحكم بأن عمر الفرامل يختلف باختلاف نوعها ، ويبدو أن النوع ب هو الأفضل .

تمارين (١٤ – ٥)

١ – أخذت عينات عشوائية من الحجم خمسة من ثلاثة أنهار كبيرة فنتجت عنها البيانات الآتية عن مستوى التلوث . استخدم طريقة كروسكال – واليس لبحث ما إذا كانت مستويات التلوث واحدة فى الأنهار الثلاثة . استخدم مستوى الدلالة ٠,١٠.

النهر الأول	النهر الثانى	النهر الثالث
٢,٧	٢,٩	٠,٦
١,٤	٢,٤	١,٢
٢,٠	٣,٧	١,٥
١,٢	١,٦	١,٧
٢,١	٢,٤	٢,١

٢ – لدراسة تأثير لون الإعلان فى جذب الزبائن قامت إحدى الشركات بتصميم خمسة عروض اعلانية متشابهة فى كل شىء ما عدا اللون المستخدم ، ثم اختارت عشرة أشخاص عشوائيا وطلب إلى كل منهم ترتيب هذه العروض بحسب مدى جاذبيتها للعين بحيث يعطى الترتيب ١ لأكثر العروض جاذبية والترتيب ٥ لأقلها جاذبية ، فجاءت النتائج كما فى الجدول الآتى . هل هناك دليل (عند المستوى ٠,١٠) على أن للون تأثير فى جذب الزبائن ؟

الأشخاص (القطاعات)	اللون السائد				
	(١) بدون ألوان	(٢) الأحمر	(٣) الأصفر	(٤) الأزرق	(٥) خليط
(١)	١	٢	٤	٣	٥
(٢)	١	٢	٥	٣	٤
(٣)	٣	٤	٥	١	٢
(٤)	١	٢	٣	٥	٤
(٥)	٤	٥	١	٣	٢
(٦)	١	٢	٣	٤	٥
(٧)	٢	١	٥	٤	٣
(٨)	١	٢	٤	٥	٣
(٩)	٣	٥	١	٢	٤
(١٠)	١	٢	٥	٤	٣

الفصل الخامس عشر

اختيار العينات وتحليلها

SELECTION AND ANALYSIS OF SAMPLES

يقدم هذا الفصل بعض طرق اختيار العينات وكيفية تحليل ما ينجم عنها من بيانات لتقدير خواص المجتمعات التى أخذت منها .

ولقد ذكرنا فى مستهل هذا الكتاب أن المجتمع الإحصائى هو مجموعة من الأشياء أو الأحداث التى تكون موضع اهتمامنا فى وقت ما من حيث متغير ما أو عدة متغيرات ، وأشرنا إلى أن دراسة مجتمع ما تقتضى أن يكون هذا المجتمع معرفا تعريفا واضحا خاصة فيما يتعلق بالمتغيرات التى ندرسها وطريقة قياسها وفى تحديد الوحدات التى يتكون منها المجتمع . ونظرا لأنه من الصعب بل قد يكون من المستحيل دراسة المجتمع بكامله فإن هذه الدراسة تقوم فى أغلب الحالات من خلال عينات تختار بحسب خطط معينة تتفق مع طبيعة المجتمع والهدف من دراسته .

والعينة لا تكون ذات قيمة إلا بالقدر الذى تمكننا به من إصدار أحكام عن الثوابت الإحصائية للمجتمع الذى أخذت منه ، ومن ثم كانت ضرورة العناية القصوى باختيار العينة التى تمثل المجتمع أفضل تمثيل ممكن وبحيث تتسم بصفات تسمح بتحقيق هذا الغرض .

(١٥ - ١) المعاينة العشوائية :

من المتطلبات الرئيسية لعملية الاستدلال الإحصائى أن تكون المعاينة من المجتمع عشوائية بمعنى أن تختار العينة بخطة تضمن عدم وجود تحيز من أى نوع قد يؤثر

فى اختيارها . ولا يغرب عن بالنا أننا حين نختار عينة عشوائية ما من حجم ما من مجتمع ما بخطة ما فإنما نكون قد اخترنا واحدة من العينات العديدة التى يمكن أن تختار بنفس الخطة وبنفس الحجم من هذا المجتمع . وإذا كانت ا هى التقدير الذى وجدناه فى إحدى العينات لأحد ثوابت المجتمع – كالوسط الحسابى – فإن ا تكون واحدة من القيم العديدة التى توجد فى العينات الأخرى والتى يمكن أن نقدر بها نفس الثابت . ولذلك نعتبر أن ا هى إحدى قيم متغير عشوائى يهمنا أن نعرف توزيع احتماله لأن هذا التوزيع هو الذى نرتكز عليه فى بناء اختبارات الدلالة وتحديد فترات الثقة وتقدير درجات الثقة فيما نصدره من قرارات عن المجتمع ، مما يدخل فى موضوع الاستدلال الاحصائى . ولقد سبق الإشارة إلى ذلك فى أكثر من مناسبة .

(١٥ – ٢) المعاينة الاحتمالية PROBABILITY SAMPLING

المعاينة الاحتمالية مصطلح عام يطلق على خطط المعاينة العشوائية التى تختار فيها العينة بحيث يكون لكل وحدة من وحدات المجتمع **احتمال معروف** للدخول فيها وبحيث تتفق طريقة الاختيار مع هذه الاحتمالات .

إن معرفة هذه الاحتمالات هى التى تتيح لنا استخدام قواعد ونظريات الاحتمال لاستنباط توزيعات الاحتمال اللازمة لعملية الاستدلال الإحصائى .

أما إذا كانت المعاينة غير عشوائية أو كانت احتمالات بعض أو كل وحدات المجتمع للدخول فى العينة لا يمكن تحديده فإن العينة تكون حينئذ غير احتمالية . وفى هذه الحال لا نستطيع استخدام الاختبارات الإحصائية أو القيام بعملية الاستدلال الاحصائى بالطرق التى مرت بنا فى الفصول السابقة . ومع ذلك لا يجوز التقليل من أهمية مثل هذه العينات التى يمكن الإفادة منها بطرق أخرى .

وهناك عدة خطط للمعاينة الاحتمالية ، نتناول منها هنا الخطط الأكثر شيوعا فى مختلف الميادين التطبيقية وهى : المعاينة العشوائية البسيطة – المعاينة الطبقية –

المعاينة متعددة المراحل – المعاينة المنتظمة – المعاينة المساحية . وتتوقف الخطة التى نختارها لدراسة مجتمع ما على طبيعة هذا المجتمع من ناحية وعلى نوع الاستنتاجات التى نريد أن نخرج بها عنه من ناحية أخرى . على أن المعيار الرئيسى الذى يجب أن نضعه نصب أعيننا فى هذا الاختيار هو الحصول على أكبر قدر ممكن من الدقة فى الحكم على المجتمع بأقل مجهود ممكن وبأقل تكلفة .

وينبغى أن تُعد خطة المعاينة بالكامل قبل القيام بالتجربة والتجميع الفعلى للبيانات وبحيث تتضمن قاعدتين : قاعدة لطريقة سحب العينة من المجتمع ، وقاعدة لتقدير ثوابت المجتمع من البيانات التى نحصل عليها من العينة مع تقدير مدى الدقة فى هذه التقديرات .

سنفترض هنا لتسهيل الدراسة أن المجتمعات منتهية وإن كان أغلب المجتمعات ذات أعداد غير منتهية من الوحدات ، وسنهتم بصفة خاصة بتقدير ثابتين هما (١) الوسط الحسابى μ لمجتمع ذى متغير كمى ، (٢) النسبة ح لمجتمع ذى متغير نوعى ذى حدين ، ويتبع كل من هذين التقديرين تقدير المجموع الكلى لقيم المتغير فى المجتمع مع العناية بتقدير مدى الثقة فى كل من هذه التقديرات .

(١٥ – ٣) العينة العشوائية البسيطة SIMPLE RANDOM SAMPLE

إن هذا النوع من العينات هو أهم أنواع العينات الاحتمالية وأبسطها ويتخذ أساسا لبناء كثير من خطط المعاينات الأخرى ، ولقد سبق أن قدمنا العينة العشوائية البسيطة بالبند (١ – ٢) حيث عرّفناها بأنها تلك العينة التى تؤخذ من المجتمع بحيث يكون لكل وحدة من وحداته احتمال **متساوى** للدخول فى العينة ، وبحيث يكون دخول أى وحدة فى العينة **مستقلا** عن الوحدات الأخرى التى قد تدخل فيها . ويمكن إثبات أن طريقة المعاينة العشوائية البسيطة للعينات التى من حجم معين ن تعطى لكل مجموعة من ن من وحدات المجتمع نفس الفرصة لتكوين عينة . وتتخذ هذه الصفة كتعريف آخر للعينة العشوائية البسيطة .

إن تعريف العينة العشوائية البسيطة يتضمن أن يكون اختيار العينة متروكا للصدفة وحدها . ولهذا فإن هذه العينة تكون مناسبة إذا كان المجتمع الذى نسحب منه متجانسا من حيث المتغير الذى نتناوله . وإذا كان المجتمع ذا حدين فإن المعاينة العشوائية البسيطة تكون مناسبة إذا كانت النسبة ح – وهى احتمال وقوع أى وحدة من وحدات المجتمع فى أحد قسمى المجتمع – واقعة بين ٢٠٪ و ٨٠٪ .

ولا نحتاج فى هذه المرحلة لأى أسس جديدة فى تناول العينات العشوائية البسيطة فقد كانت هى التى نتناولها طوال دراستنا فى الفصول السابقة . على أنه من المهم أن نتذكر دائما أنه إذا أخذت عينة عشوائية بسيطة من مجتمع متوسطه μ وتباينه σ^2 فإن توزيع المعاينة للأوساط الحسابية للعينات التى من الحجم ن يكون متوسطه $\mu_{\bar{س}} = \mu$ وتباينه $\sigma^2_{\bar{س}} = \frac{\sigma^2}{ن}$ راجع البند (٦ – ٢) – وإذا كان المجتمع معتدلا فإن توزيع المعاينة هذا يكون معتدلا : مع $(\mu ، \frac{\sigma}{\sqrt{ن}})$ ، وإذا لم يكن المجتمع معتدلا وكان حجم العينة كبيرا فإن توزيع المعاينة يقترب من هذا التوزيع المعتدل كلما زاد حجم العينة – راجع البند (٦ – ٣) . وهذه الحقيقة تيسر لنا التوصل إلى الاستنتاجات التى تتعلق بمتوسطات المجتمعات ومجاميعها .

(١٥ – ٣ – ١) تقدير الوسط الحسابى والمجموع :

اعتبر مجتمعا حجمه ه ومتوسطه μ وتباينه σ^2 . إن المجموع الكلى لوحدات هذا المجتمع هو م = ه μ . نريد تقدير كل من μ ، م من عينة عشوائية بسيطة $س_1$ ، $س_2$ ، . . . ، $س_ن$ حجمها ن مأخوذة من هذا المجتمع . نعلم ما يلى :

(أولا) إذا كان متوسط العينة $\bar{س}$ حيث $\bar{س} = \frac{1}{ن} \sum_{ر=1}^{ن} س_ر$

فإن هذا المتوسط هو تقدير غير متحيز للمتوسط μ للمجتمع . وبالتالى فإن

$$\hat{م} = ه \, \bar{س} \qquad (1)$$

هو تقدير غير متحيز للمجموع م للمجتمع .

(ثانيا) إذا كان تباين العينة $ع^2$ حيث $ع^2 = \frac{1}{ن - 1} \sum (س_ر - \bar{س})^2$

فإن $ع^2 \left(1 - \frac{ن}{هـ}\right)$ يكون تقديرا غير متحيز للتباين σ^2 للمجتمع ، ومن هذا نستطيع اثبات أن $\frac{ع^2}{ن}\left(1 - \frac{ن}{هـ}\right)$ ، $\frac{هـ^2 ع^2}{ن}\left(1 - \frac{ن}{هـ}\right)$ هما على الترتيب تقديران غير متحيزين للتباين لتوزيع المعاينة للمتوسطات وتوزيع المعاينة للمجاميع للعينات ذوات الحجم ن .

ويقدر الخطأ المعيارى للمتوسطات بالمقدار $ع_{\bar{س}} = \frac{ع}{\sqrt{ن}} \sqrt{1 - \frac{ن}{هـ}}$ (٢)

كما يقدر الخطأ المعيارى للمجاميع بالمقدار $ع_م = \frac{هـ ع}{\sqrt{ن}} \sqrt{1 - \frac{ن}{هـ}}$ (٣)

وهذان التقديران متحيزان تحيزا قليلا ولكننا نتجاوز عن ذلك فى معظم التطبيقات .

يلاحظ أنه إذا كان هناك تقدير غير متحيز لتباين توزيع ما فإن جذره التربيعى ليس من الضرورى أن يكون تقديرا غير متحيز للانحراف المعيارى للتوزيع .)

ويعرف العامل $\sqrt{1 - \frac{ن}{هـ}}$ أو مربعه بأنه عامل التصحيح للمجتمعات المنتهية Finite population correction factor ويمكن إهماله إذا كانت النسبة $\frac{ن}{هـ}$ (وهى نسبة حجم العينة إلى حجم المجتمع) تقل عن حوالى ١٠٪ لأن عامل التصحيح يكون فى هذه الحال قريبا من الواحد الصحيح . ويلاحظ من (٢) و(٣) أن كلا من الخطأين المعيارين $ع_{\bar{س}}$ ، $ع_م$ يساوى صفرا إذا كان ن = هـ وهذا ما يجب أن يكون لأننا فى هذه الحال نكون قد استخدمنا جميع وحدات المجتمع ولا مجال للحديث عن توزيعات المعاينة أو الأخطاء المعيارية .

مثال (١٥ – ١) :

جمعت توقيعات على التماس ما فى ٦٧٦ بطاقة ، وكانت كل بطاقة قد أعدت لتكفى ٤٢ توقيعا، غير أن بعض البطاقات اشتملت على عدد من التوقيعات يقل

عن ٤٢ . أخذت عينة عشوائية بسيطة من ٥٠ بطاقة (أى بواقع حوالى ٧,٤٪) وحسب عدد التوقيعات بكل منها ووضعت النتيجة فى التوزيع التكرارى المبين بالجدول (١٥ — ١) .

الجدول (١٥ – ١)

عدد التوقيعات $س_ر$	التكرار $ك_ر$	عدد التوقيعات $س_ر$	التكرار $ك_ر$
٤٢	٢٣	١٤	١
٤١	٤	١١	١
٣٦	١	١٠	١
٣٢	١	٩	١
٢٩	١	٧	١
٢٧	٢	٦	٣
٢٣	١	٥	٢
١٩	١	٤	١
١٦	٢	٣	١
١٥	٢		
			٥٠

مجـ ك س = ١٤٧١ ، مجـ ك $س^٢$ = ٥٤٤٩٧

(أولا) أوجد تقديرا للعدد الكلى للتوقيعات على هذا الالتماس .

(ثانيا) أوجد فترة ثقة بدرجة ٨٠٪ للعدد الكلى للتوقيعات .

الحل :

حجم المجتمع هـ = ٦٧٦ بطاقة ، حجم العينة ن = ٥٠ بطاقة

(أولا) $\bar{س} = \frac{١}{ن}$ مجـ ك س = $\frac{١}{٥٠}$ × ١٤٧١ = ٢٩,٤٢

وهذا هو الوسط الحسابى لعدد التوقيعات فى العينة .

من (١) ، نقدر العدد الكلى للتوقيعات بالمقدار

م = هـ $\bar{س}$ = ٦٧٦ × ٢٩,٤٢ = ١٩٨٨٨ توقيعا .

(ثانيا) لإيجاد فترة الثقة المطلوبة نحتاج إلى إيجاد الخطأ المعيارى للمجاميع وهذا بدوره يحتاج إلى إيجاد تباين العينة ع^{٢} كالآتى .

$$ع^{٢} = \frac{١}{ن - ١} \left[مجـ ك س^{٢} - \frac{(مجـ ك س)^{٢}}{ن} \right]$$

$$= \frac{١}{٤٩} \left(٥٤٤٩٧ - \frac{(١٤٧١)^{٢}}{٥٠} \right) = ٢٢٨,٩٨$$

∴ ع = ١٥,١٣

من (٣) ، $ع_{م} = \frac{هـ ع}{\sqrt{ن}} \sqrt{١ - \frac{ن}{هـ}} = \frac{٦٧٦ \times ١٥,١٣}{\sqrt{٥٠}} \sqrt{١ - \frac{٥٠}{٦٧٦}}$

$$= ١٣٩١,٤٨$$

بالرغم من أن التوزيع الأصلى لعدد التوقيعات يبدو بعيدا عن الاعتدال إلا أننا نستطيع أن نعتبر أن توزيع المعاينة للمتوسطات ، وبالتالى للمجاميع ، هو توزيع معتدل على وجه التقريب لأن حجم العينة كبيرا (ن = ٥٠) . ومن جدول المساحات أسفل المنحنى المعتدل المعيارى نجد أن القيمتين الحرجتين اللتين يقع بينهما ٨٠٪ من التوزيع هما ± ١,٢٨ . وكما فى البند (٦ – ٦ – ٢) أو البند (٦ – ١٠ – أولا) نجد ما يلى :

الحد الأدنى لمجموع التوقيعات = ١٩٨٨٨ – ١٣٩١,٤٨ × ١,٢٨ = ١٨١٠٧

الحد الأعلى لمجموع التوقيعات = ١٩٨٨٨ + ١٣٩١,٤٨ × ١,٢٨ = ٢١٦٦٩

وإذن الفترة (١٨١٠٧ ، ٢١٦٦٩) هى فترة ثقة بدرجة ٨٠٪ لعدد التوقيعات على الالتماس .

(أظهر العد الكلى لجميع البطاقات أن عدد التوقيعات ٢١٠٤٥)

(١٥ – ٣ – ٢) تقدير النسبة ح ومجموع الوحدات :

فى المجتمع ذى الحدين تكون كل وحدة من وحدات المجتمع منتمية إلى واحد من اثنين من الأقسام ا ، اَ وينصب اهتمامنا على تقدير الدليل ح وهو نسبة الوحدات التى تقع فى أحد القسمين وليكن القسم ا وعلى تقدير المجموع الكلى للوحدات فى هذا القسم .

نفرض أننا أخذنا من هذا المجتمع عينة عشوائية بسيطة حجمها ن . نذكر أنه فى توزيع ذى الحدين للعينات التى من الحجم ن يكون للمتغير س وسط حسابى ن ح وتباين ن ح (١ – ح) كما يكون لنسبة هذا المتغير وسط حسابى ح وتباين $\frac{ح (١ - ح)}{ن}$ وعلى ذلك فإن تناول هذه الحالة يمكن أن يتخذ نفس طريقة تناول المجتمع الكمى مع وضع ح بدلا من μ ، ح (١ – ح) بدلا من $\sigma^٢$.

وينتج ما يلى :

(أولا) النسبة $ر = \frac{س}{ن}$ (٤)

وهى نسبة عدد وحدات العينة التى تنتمى إلى القسم ا إلى العدد الكلى للوحدات فى العينة ، هى تقدير غير متحيز للنسبة ح .

كما أن العدد $م = هـ ر$ (٥)

هو تقدير غير متحيز لمجموع الوحدات الواقعة فى القسم ا فى المجتمع .

(ثانيا) المقدار $ع_ر = \sqrt{\frac{ر(١ - ر)}{ن}} \sqrt{١ - \frac{ن}{هـ}}$ (٦)

هو تقدير للخطأ المعيارى للنسبة ر

والمقدار $ع_م = هـ ع_ر$ (٧)

هو تقدير للخطأ المعيارى لمجموع الوحدات فى القسم ا

والصيغ (٥) ، (٦) ، (٧) هى نفس الصيغ (١) ، (٢) ، (٣) بعد وضع ر بدلا من $\bar{س}$ ووضع ر (١ - ر) بدلا من $ع^٢$.

مثال (١٥ - ٢) :

فى قائمة من ٣٠٤٢ اسما وعنوانا سحبت عينة عشوائية بسيطة من ٢٠٠ اسم فظهر فيها أن هناك خطأ فى ٣٨ عنوانا . قدر العدد الكلى للعناوين التى تحتاج إلى تصحيح وأوجد الخطأ المعيارى لهذا التقدير .

الحل :

حجم المجتمع هـ = ٣٠٤٢ شخصا وحجم العينة ن = ٢٠٠ شخصا

$ر = \frac{٣٨}{٢٠٠} = ٠,١٩$ (نسبة العناوين الخاطئة فى العينة)

من (٥) ، نقدر المجموع الكلى للعناوين الخاطئة بالمقدار

$م = هـ ر = ٣٠٤٢ \times ٠,١٩ = ٥٧٨$ عنوانا خاطئاً .

من (٧) ، الخطأ المعيارى لهذا المجموع هو :

$$ع_م = ٣٠٤٢ \sqrt{\frac{٠,٨١ \times ٠,١٩}{٢٠٠}} = ٨٤,٣٨$$

وقد أهملنا عامل التصحيح لأن $\frac{ن}{هـ} = \frac{٢٠٠}{٣٠٤٢}$ = ٦,٥٧٪ وهى نسبة صغيرة .

(١٥ – ٣ – ٣) حجم العينة :

كما سبق القول مرارا ، كلما كبر حجم العينة كلما زادت ثقتنا فيما نستخلصه من نتائج . ولذلك ينبغى أن نحرص على ألا يكون حجم العينة صغيرا بدرجة تكون معها دقة تقديراتنا أقل مما يجب . غير أنه ينبغى فى الوقت نفسه أن نتجنب أن يكون حجم العينة كبيرا بدرجة تثقل كاهلنا بالجهد والتكاليف . وبالتالى فإن الخطوة الأولى فى عملية التجريب هى تحديد الحجم المناسب للعينة . وفى هذا الصدد نحيل القارىء إلى البند (٦ – ١٠) وبصفة خاصة إلى الصيغة (٢٣) التى تعطينا الحد الأعلى لحجم العينة عندما تكون المعاينة من مجتمع معتدل ، وهى :

$$ن = \left(\frac{مع_{\frac{\alpha}{٢}}\ \sigma}{خ}\right)^٢ \qquad (٨)$$

والصيغتين (٢٦) ، (٢٧) فى حالة المعاينة من مجتمع ذى حدين وهما :

$$ن = ح ك \left(\frac{مع_{\frac{\alpha}{٢}}}{خ}\right)^٢ \qquad \text{حيث } ك = ١ - ح \qquad (٩)$$

$$، ن = \frac{١}{٤}\left(\frac{مع_{\frac{\alpha}{٢}}}{خ}\right)^٢ \qquad (١٠)$$

في كثير من الأحيان يعرف الباحث أو يكون لديه ما يدعو إلى الشك فى أن المجتمع غير متجانس من حيث المتغير الذى يدرسه ، بل ينقسم إلى عدد من القطاعات تختلف الاستجابات فيها بين كل قطاع وآخر بينما تتجانس داخل كل قطاع على حدة . وإذا كان الأمر كذلك نقول إن المجتمع مقسم إلى طبقات أو شرائح تحددها تركيبة المجتمع ، وهذه الطبقات قد تكون بحسب الجنس أو العمر أو الجنسية أو المستوى الثقافى أو درجة الإصابة بمرض ما . فى هذه الحال لا تكون العينة العشوائية البسيطة صالحة لتمثيل المجتمع ، بل تكون خطة المعاينة المناسبة هى تلك المسماة بالمعاينة العشوائية الطبقية البسيطة ، أو اختصارا بالمعاينة الطبقية . وتتلخص هذه الخطة فى تحديد طبقات المجتمع بحيث لا تتداخل طبقة مع أخرى ثم أخذ عينة عشوائية بسيطة من كل طبقة بحيث تكون المعاينة مستقلة من طبقة إلى أخرى . وتتألف العينة المطلوبة من مجموع هذه العينات الجزئية .

وتبدأ الخطة بتحديد الحجم الكلى للعينة أو النسبة التى يرى أخذها من الحجم الكلى للمجتمع ، ثم تحديد أحجام العينات الجزئية مع الأخذ فى الاعتبار أحجام الطبقات والتباين داخل كل طبقة ، أو أى عوامل أخرى تؤثر فى تركيب المجتمع .

مثال (١٥ – ٣) :

نفرض أن لدينا مجتمعا حجمه ٤٠٠٠ وأن الإمكانات لا تسمح إلا بفحص عينة حجمها ٦٠ أى بنسبة $\frac{٦٠}{٤٠٠٠}$ = ٠,٠١٥ من حجم المجتمع . ونفرض أننا نعرف أن المجتمع مقسم إلى ثلاث طبقات أحجامها ٢٠٠٠ ، ١٢٠٠ ، ٨٠٠ . نظرا لاختلاف أحجام الطبقات فإن العينة تكون أقدر تمثيلا للمجتمع إذا أتحنا للطبقة ذات الحجم الأكبر أن تسهم بقدر أكبر فى العينة ، وللطبقة ذات الحجم الأصغر أن تسهم بقدر أقل . ولتحقيق هذه العدالة نستخدم الطريقة الآتية .

طريقة التقسيم المتناسب

PROPORTIONAL ALLOCATION METHOD

تنص هذه الطريقة على أن نأخذ من كل طبقة عينة عشوائية بسيطة يتناسب حجمها مع حجم الطبقة . فإذا رمزنا لحجم المجتمع بالرمز ﻫ وللحجم الكلى للعينة بالرمز ن وكان المجتمع مقسما إلى ك من الطبقات أحجامها $ﻫ_{١}$ ، $ﻫ_{٢}$ ، . . . ، $ﻫ_{ك}$ (مجموعها ﻫ) فإننا نأخذ من هذه الطبقات عينات أحجامها $ن_{١}$ ، $ن_{٢}$ ، . . . ، $ن_{ك}$ (مجموعها ن) بحيث

$$\frac{ن_{١}}{ﻫ_{١}} = \frac{ن_{٢}}{ﻫ_{٢}} = \ldots\ldots = \frac{ن_{ك}}{ﻫ_{ك}}$$

مع ملاحظة أن كلا من هذه النسب يساوى $\frac{ن}{ﻫ}$ وهو فى هذا المثال يساوى ٠,٠١٥ . ومن السهل أن نرى أن الحجم $ن_{و}$ الذى يؤخذ من الطبقة و يكون على الصورة

$$ن_{و} = \frac{ن}{ﻫ} \times ﻫ_{و} \quad ، \quad و = ١ ، ٢ ، \ldots ، ك \qquad (١١)$$

ففى المثال (١٥ – ٣) تكون أحجام العينات الجزئية كما يلى – انظر الجدول (١٥ – ٢) :

$$ن_{١} = ٠,٠١٥ \times ٢٠٠٠ = ٣٠$$

$$ن_{٢} = ٠,٠١٥ \times ١٢٠٠ = ١٨ ،$$

$$ن_{٣} = ٠,٠١٥ \times ٨٠٠ = ١٢ ،$$

الجدول (١٥ - ٢)

أحجام العينات بطريقة التقسيم المتناسب

الطبقة	حجم الطبقة $هـ_{ى}$	حجم العينة $ن_{ى}$
(١)	٢٠٠٠	٣٠
(٢)	١٢٠٠	١٨
(٣)	٨٠٠	١٢
المجموع	٤٠٠٠	٦٠

الجدول (١٥-٣)

أحجام العينات بطريقة التقسيم الأمثل

الطبقة	حجم الطبقة $هـ_{ى}$	الانحراف المعيارى $\sigma_{ى}$	حجم العينة $ن_{ى}$
(١)	٢٠٠٠	٤	٣١
(٢)	١٢٠٠	٣	١٤
(٣)	٨٠٠	٥	١٥
المجموع	٤٠٠٠		٦٠

إن المعاينة بطريقة التقسيم المتناسب تأخذ فى الاعتبار الفروق بين أحجام الطبقات ولا تأخذ فى الاعتبار الفروق بين التباينات داخل هذه الطبقات بل تعتبر أن هذه التباينات متساوية . وإذا كانت التباينات تختلف من طبقة لأخرى فمن الأفضل أن نأخذ عينات أكبر حجما من الطبقات الأكثر تشتتا وعينات أصغر حجما من الطبقات الأقل تشتتا ، ولتحقيق ذلك نستخدم الطريقة الآتية .

طريقة التقسيم الأمثل **OPTIMUM ALLOCATION METHOD**

تنص هذه الطريقة على أن نأخذ من كل طبقة عينة عشوائية بسيطة يتناسب حجمها مع كل من حجم الطبقة وانحرافها المعيارى ، فإذا كان σ_1 ، σ_2 ، ... ، $\sigma_{\text{ك}}$ ترمز إلى الانحرافات المعيارية للطبقات فإن الأحجام التى تؤخذ من هذه الطبقات تكون بحيث :

$$\frac{\text{ن}_1}{\text{هـ}_1\,\sigma_1} = \frac{\text{ن}_2}{\text{هـ}_2\,\sigma_2} = \dots = \frac{\text{ن}_{\text{ك}}}{\text{هـ}_{\text{ك}}\,\sigma_{\text{ك}}}$$

ويمكن اثبات أن هذه المتساويات تؤدى إلى أن يكون الحجم $\text{ن}_{\text{ه}}$ الذى يؤخذ من الطبقة ه على الصورة الآتية :

$$\text{ن}_{\text{ه}} = \frac{\text{ن}}{\text{هـ}_1\sigma_1 + \text{هـ}_2\sigma_2 + \dots + \text{هـ}_{\text{ك}}\sigma_{\text{ك}}} \times \text{هـ}_{\text{ه}}\,\sigma_{\text{ه}} \;:\; \text{ه} = 1 ، 2 ، \dots ، \text{ك}$$

$$= \frac{\text{ن}}{\sum_{\text{ر}=1}^{\text{ك}} \text{هـ}_{\text{ر}}\,\sigma_{\text{ر}}} \times \text{هـ}_{\text{ه}}\,\sigma_{\text{ه}} \qquad (١٢)$$

ويلاحظ أنه إذا كانت الانحرافات المعيارية للطبقات متساوية جميعها فإن الصغية (١٢) تؤول بالضبط إلى الصيغة (١١) .

وفى المثال إذا كان $\sigma_١ = ٤$ ، $\sigma_٢ = ٣$ ، $\sigma_٣ = ٥$ فإن أحجام العينات الجزئية تحسب كما يلى : انظر الجدول (١٥ - ٣) .

$$\text{مج } ﻫ_ر \sigma_ر = ٢٠٠٠ \times ٤ + ١٢٠٠ \times ٣ + ٨٠٠ \times ٥ = ١٥٦٠٠$$

$$\therefore ن_١ = \frac{٦٠}{١٥٦٠٠} \times ٢٠٠٠ \times ٤ = ٣١$$

$$، ن_٢ = \frac{٦٠}{١٥٦٠٠} \times ١٢٠٠ \times ٣ = ١٤$$

$$، ن_٣ = \frac{٦٠}{١٥٦٠٠} \times ٨٠٠ \times ٥ = ١٥$$

هذا مع ملاحظة أنه عند التعويض فى أى من الصيغتين (١١) أو (١٢) نأخذ أقرب عدد صحيح للقيمة التى تنتج من هذا التعويض . وفى الصيغة (١٢) إذا كانت الانحرافات المعيارية للطبقات غير معروفة فينبغى تقديرها من عينات سابقة .

تقدير البارامترات :

بعد تحديد أحجام العينات سواء بطريقة التقسيم المتناسب أو التقسيم الأمثل نقوم بسحب العينات من الطبقات بحسب هذه الأحجام ثم نجرى ما نريد من قياسات كقياس الطول أو الوزن ... على وحدات هذه العينات لنحصل على مجموعة من القيم لكل عينة . من هذه القيم نحسب متوسطات العينات $\bar{س}_١$ ، $\bar{س}_٢$ ، ... ، $\bar{س}_ل$ وانحرافاتها المعيارية $ع_١$ ، $ع_٢$ ، .... ، $ع_ل$ وذلك لاستخدامها فيما يلى :

(أولا) تقدير متوسطات الطبقات :

نظرا لأن العينات المسحوبة هى عينات عشوائية بسيطة فإن متوسط الطبقة ومجموعها والخطأ المعيارى لتوزيع المعاينة تقدر بنفس الصيغ المبينة بالبند (١٥ - ٣ - ١) السابق .

(ثانيا) تقدير متوسط المجتمع والخطأ المعيارى .

(أ) تقدير الوسط الحسابى

يقدر الوسط الحسابى μ للمجتمع تقديرا غير متحيز بالمقدار $\bar{س}$ حيث

$$\bar{س} = \frac{1}{هـ} (هـ_1 \bar{س}_1 + هـ_2 \bar{س}_2 + \dots + هـ_ك \bar{س}_ك)$$

$$= \frac{1}{هـ} \sum_{1}^{ك} هـ_ى \bar{س}_ى \qquad (١٣)$$

مع ملاحظة أن متوسطات العينات $\bar{س}_ى$ قد رجحت بأحجام الطبقات وليس بأحجام العينات .

وحين تكون المعاينة بطريقة التقسيم المتناسب فإن $\frac{هـ_ى}{هـ} = \frac{ن_ى}{ن}$

وذلك من الصيغة (١١) وفى هذه الحالة تؤول الصيغة (١٣) إلى الصيغة الآتية :

$$\bar{س} = \frac{1}{ن} \sum_{1}^{ك} ن_ى \bar{س}_ى \qquad (١٤)$$

أى أن الوسط الحسابى μ للمجتمع يقدر فى هذه الحالة بواسطة الوسط الحسابى للمشاهدات فى العينة الكلية التى تنتج من ضم العينات الجزئية معا .

وفى كلتا الحالتين يقدر المجموع الكلى للمجتمع بالمقدار .

$$م = هـ \bar{س} \qquad (١٥)$$

(ب) تقدير الخطأ المعيارى .

من الصيغة (١٣) يمكن إثبات أن التباين $\sigma^2_{\bar{س}}$ لتوزيع المعاينة للوسط الحسابى يقدر بدلالة تباينات العينات $ع^2_ى$ بالمقدار $ع^2_ع$ حيث

$$ع^2_ع = \sum و^2_ى \times \frac{ع^2_ى}{ن_ى} (1 - \frac{ن_ى}{هـ_ى}) \qquad (١٦)$$

وحيث $\text{و}_{\text{ى}} = \frac{\text{هـ}_{\text{ى}}}{\text{هـ}}$ = نسبة حجم الطبقة ى إلى الحجم الكلى للمجتمع .

أما الخطأ المعيارى للمتوسطات فيقدر بالجذر التربيعى لهذا المقدار .

وإذا كانت المعاينة الطبقية بالتقسيم المتناسب للأحجام فإن عوامل التصحيح تكون واحدة لجميع الطبقات وكل منها يساوى $1 - \frac{\text{ن}}{\text{هـ}}$ كما أن

$$\text{ن}_{\text{ى}} = \frac{\text{هـ}_{\text{ى}}}{\text{هـ}} \times \text{ن}$$

وبالتعويض بهذا المقدار فى (١٦) تؤول إلى الصيغة الآتية

$$\text{ع}^{2}_{\text{م}} = \frac{1}{\text{ن}\ \text{هـ}} \left(1 - \frac{\text{ن}}{\text{هـ}}\right) \text{مجـ}\ \text{هـ}_{\text{ى}}\ \text{ع}^{2}_{\text{ى}} \qquad (١٧)$$

وإذا أمكن أن نعتبر أن تباينات الطبقات $\text{ع}^{2}_{\text{ى}}$ متساوية وكل منها يساوى $\text{ع}^{2}_{\text{د}}$ فإن الصيغة (١٧) لحالة التقسيم المتناسب تؤول إلى الصيغة البسيطة الآتية :

$$\text{ع}^{2}_{\text{م}} = \frac{\text{ع}^{2}_{\text{د}}}{\text{ن}} \left(1 - \frac{\text{ن}}{\text{هـ}}\right) \qquad (١٨)$$

والجذر التربيعى لهذه الصيغة هو بالضبط الصيغة (٢) لتقدير الخطأ المعيارى فى حالة المعاينة العشوائية البسيطة من مجتمع منتهى فيما عدا أن التباين المشترك $\text{ع}^{2}_{\text{د}}$ يحسب هنا من داخل العينات الجزئية كالآتى :

$$\text{ع}^{2}_{\text{د}} = \frac{1}{\text{ن} - \text{ك}} \left[(\text{ن}_{1} - 1)\,\text{ع}^{2}_{1} + (\text{ن}_{2} - 1)\,\text{ع}^{2}_{2} + \ldots + (\text{ن}_{\text{ك}} - 1)\,\text{ع}^{2}_{\text{ك}} \right] \qquad (١٩)$$

وفى جميع الحالات تقدر تباينات المجاميع من تباينات المتوسطات بالضرب فى مربع حجم المجتمع وهو هـ^{2} .

مثال (١٥ – ٤) :

فى المثال (١٥ – ٣) إذا كانت أحجام العينات قد حسبت بطريقة التقسيم المتناسب ، وعلى فرض تساوى تباينات الطبقات ، فاوجد فترة ثقة بدرجة ٩٥٪ لمتوسط المجتمع علما بأن الأوساط الحسابية والانحرافات المعيارية للعينات كالآتى :

$\bar{س}_١ = ٧$ ، $\bar{س}_٢ = ١٠$ ، $\bar{س}_٣ = ١٣$ ، $ع_١ = ٣$ ، $ع_٢ = ٢,٥$ ، $ع_٣ = ٤$.

الحل :

من الجدول (١٥ – ٢) والصيغة (١٤) نجد أن

$$\bar{س} = \frac{١}{ن} \sum ن_ى \bar{س}_ى$$

$$= \frac{١}{٦٠} (٣٠ \times ٧ + ١٨ \times ١٠ + ١٢ \times ١٣) = ٩,١$$

من الصيغة (١٩) نقدر التباين المشترك للطبقات كالآتى :

$$ع_د^٢ = \frac{١}{٦٠ - ٣} (٢٩ \times ٩ + ١٧ \times ٦,٢٥ + ١١ \times ١٦) = ٩,٥٣$$

$$\therefore ع_د = ٣,٠٩$$

من الصيغة (١٨) نقدر الخطأ المعيارى لتوزيع المعاينة للأوساط الحسابية كالآتى مع ملاحظة أن عامل التصحيح قريب من الواحد ويمكن اهماله :

$$ع_{\bar{س}} = \frac{ع_د}{\sqrt{ن}} = \frac{٣,٠٩}{\sqrt{٦٠}} = ٠,٤٠$$

ولما كان حجم العينة كبيرا (ن = ٦٠) يمكن أن نعتبر أن توزيع المعاينة معتدلا ويكون حدا الثقة بدرجة ٩٥٪ لمتوسط المجتمع على الصورة الآتية :

$$\bar{س} \pm ع_{خ} \times ١,٩٦$$

∴ الحد الأدنى للفترة = ٩,١ − ٠,٤٠ × ١,٩٦ = ٨,٣٢

والحد الأعلى للفترة = ٩,١ + ٠,٤٠ × ١,٩٦ = ٩,٨٨

وبذلك تكون الفترة (٨,٣٢ ، ٩,٨٨) هى فترة ثقة بدرجة ٩٥٪ لمتوسط المجتمع .

مثال (١٥ − ٥) :

فى المثال (١٥ − ١) إذا كانت أحجام العينات قد حسبت بطريقة التقسيم الأمثل فاوجد فترة ثقة بدرجة ٩٩٪ لمتوسط المجتمع علما بأن الأوساط الحسابية للعينات هى $\bar{س}_{١}$ = ٨ ، $\bar{س}_{٢}$ = ١٢ ، $\bar{س}_{٣}$ = ١٣ وأن الانحرافات المعيارية للطبقات هى كما جاءت بالجدول (١٥ − ٣) : $\sigma_{١}$ = ٤ ، $\sigma_{٢}$ = ٣ ، $\sigma_{٣}$ = ٥ .

الحل :

من الجدول (١٥ − ٣) والصيغة (١٣) نجد أن

$$\bar{س} = \frac{١}{هـ} \text{مج } هـ_{ى} \bar{س}_{ى}$$

$$= \frac{١}{٤٠٠٠} (٨ \times ٢٠٠٠ + ١٢ \times ١٢٠٠ + ١٣ \times ٨٠٠) = ١٠,٢$$

من الصيغة (١٦) نقدر الخطأ المعيارى لتوزيع المعاينة للأوساط الحسابية كالآتى ، مع إهمال عوامل التصحيح لقرب كل منها من الواحد الصحيح .

لدينا : $و_{١}^{٢} = (\frac{٢٠٠٠}{٤٠٠٠})^{٢} = ٠,٢٥$ ، $و_{٢}^{٢} = (\frac{١٢٠٠}{٤٠٠٠})^{٢} = ٠,٠٩$

، $و_{م}^{2} = (\frac{٨٠٠}{٤٠٠٠})^{2} = ٠,٠٤$

$\therefore ع_{غ}^{2} = \frac{٠,٢٥ \times ١٦}{٣١} + \frac{٠,٠٩ \times ٩}{١٤} + \frac{٠,٠٤ \times ٢٥}{١٥} = ٠,٢٥٤$

$\therefore ع_{غ} = ٠,٥٠$

ويكون حدا الثقة بدرجة ٩٩٪ لمتوسط المجتمع على الصورة

$$\bar{س} \pm ع_{غ} \times ٢,٥٨$$

وبالتعويض فى هذه الصيغة نجد أن الفترة (٨,٩١ ، ١١,٤٩) هى فترة ثقة بدرجة ٩٩٪ لمتوسط المجتمع .

مثال (١٥ – ٦) :

حسب تعداد السكان فى سنة ما فى ٦٤ مدينة من مدن إحدى الدول . وقد قسمت هذه المدن إلى طبقتين تتألف الأولى من المدن الأكبر حجما وعددها ١٦ مدينة وتتألف الثانية من الـ ٤٨ مدينة الباقية ، ولخصت البيانات فى الجدول الآتى .

الجدول (١٥ – ٤)

الطبقة	الحجم $ه_{ر}$	$مج س_{ر}$	$مج س_{ر}^{2}$
(١)	١٦	١٠٠٧٠	٧١٤٥٥٤٠
(٢)	٤٨	٩٤٩٨	٢١٤١٧٢٠

إذا قُدر المجموع الكلى للسكان فى تلك السنة من عينة من ٢٤ مدينة فأوجد الخطأ المعيارى لهذا التقدير

(أولا) إذا كانت العينة عشوائية بسيطة ،

(ثانيا) إذا كانت العينة طبقية ذات تقسيم متناسب ،

(ثالثا) إذا كانت العينة طبقية وأخذت ١٢ مدينة من كل طبقة .

الحل :

من البيانات المعطاة نستطيع أن نحسب تباين المجتمع وتباينات الطبقات ولا حاجة لنا إذن لتقدير هذه التباينات من العينة .

(أولا) إذا كانت العينة عشوائية بسيطة فإنها تكون قد أخذت من المجتمع ككل بصرف النظر عن الطبقات . ويكون لدينا ما يلى :

حجم المجتمع هـ = ٦٤ مدينة ، حجم العينة ن = ٢٤ مدينة

مجـ س = ١٠٠٧٠ + ٩٤٩٨ = ١٩٥٦٨ ألف نسمة (مجموع السكان)

مجـ س^2 = ٧١٤٥٥٤٠ + ٢١٤١٧٢٠ = ٩٢٨٧١٧٠

$\therefore$ تباين المجتمع = σ^2 = $\frac{1}{٦٤}$ [٩٢٨٧١٧٠ − $\frac{(١٩٥٦٨)^2}{٦٤}$] = ٥١٦٢٨,٩٧

$\therefore$ σ = ٢٢٧,٢٢

من الصيغة (٣) : ع$_{م}$ = $\frac{\sigma\ هـ}{\sqrt{ن}}$ $\sqrt{١ - \frac{ن}{هـ}}$ = $\frac{٢٢٧,٢٢ \times ٦٤}{\sqrt{٢٤}}$ $\sqrt{١ - \frac{٢٤}{٦٤}}$

= ٢٣٤٦,٧٢ (الخطأُ المعيارى لمجموع السكان)

(ثانيا) إذا كانت المعاينة طبقية وبالتقسيم المتناسب فإن حجمى العينتين يكونان كالآتى :

$ن_1 = \frac{ن}{هـ} \times هـ_1 = \frac{٢٤}{٦٤} \times ١٦ = ٦$ ، $ن_2 = \frac{٢٤}{٦٤} \times ٤٨ = ١٨$

تباين الطبقة الأولى $ع_1^2 = \frac{١}{١٦} [٧١٤٥٤٥٠ - \frac{(١٠٠٧٠)^2}{١٦}] = ٥٠٤٧٧,٧٤$

تباين الطبقة الثانية $ع_2^2 = \frac{١}{٤٨} [٢١٤١٧٢٠ - \frac{(٩٤٩٨)^2}{٤٨}] = ٥٤٦٤,٦٥$

نلاحظ أن تباين الطبقة الأولى حوالى عشرة أمثال تباين الطبقة الثانية ولذلك لا نستطيع اعتبارهما متساويين .

بضرب الصيغة (١٧) فى مربع حجم المجتمع ينتج أن تباين مجموع المجتمع هو

$$ع_خ^2 = \frac{٦٤^2}{٢٤ \times ٦٤} (١ - \frac{٢٤}{٦٤}) [١٦ \times ٥٠٤٧٧,٧٤ + ٤٨ \times ٥٤٦٤,٦٥]$$

$$= \frac{٤٠}{٢٤} \times ١٠٦٩٩٤٧ = ١٧٨٣٢٤٥$$

$$\therefore ع_خ = ١٣٣٥,٣٨$$

(ثالثا) إذا كانت المعاينة طبقية وأخذنا $ن_1 = ن_2 = ١٢$ نستخدم الصيغة العامة (١٦) بعد ضربها فى مربع حجم العينة مع ملاحظة ما يلى :

$و_1 = \frac{١٦}{٦٤}$ ، $و_2 = \frac{٤٨}{٦٤}$ ، $ن_1 = ١٢$ ، $ن_2 = ١٢$

، $(١ - \frac{١٢}{١٦}) = \frac{٤}{١٦}$ ، $١ - \frac{١٢}{٤٨} = \frac{٣٦}{٤٨}$

$$ع_خ^2 = \frac{٦٤^2}{١٢} [\frac{١٦^2}{٦٤^2} \times ٥٠٤٧٧,٧٤ \times \frac{٤}{١٦} + \frac{٤٨^2}{٦٤^2} \times ٥٤٦٤,٦٥ \times \frac{٣٦}{٤٨}]$$

$$= \frac{١}{١٢} (١٦ \times ٥٠٤٧٧,٧٤ \times ٤ + ٤٨ \times ٥٤٦٤,٦٥ \times ٣٦)$$

$$= ١٠٥٦١١٢٤,١$$

$$\therefore \text{ع}_{\text{خ}} = ١٠٢٧,٦٨$$

فى هذا المثال ، بمقارنة الأخطاء المعيارية وهى ٢٣٤٦,٧٢ ، ١٣٣٥,٣٨ ، ١٠٢٧,٦٨ نجد أن أخذ حجمين متساويين للعينتين كان أكثر دقة من طريقة التقسيم المتناسب وكلاهما أدق كثيرا من طريقة المعاينة العشوائية البسيطة .

المعاينة الطبقية من مجتمع ذى حدين :

نستخدم نفس الصيغ التى قدمت فى حالة المعاينة الطبقية من مجتمعات كمية مع وضع النسب $\text{س}_{١}$ ، $\text{س}_{٢}$ ، .... ، $\text{س}_{\text{ك}}$ المحسوبة من العينات بدلا من المتوسطات الحسابية ووضع التباينات $\text{س}_{\text{ى}} (١ - \text{س}_{\text{ى}})$ بدلا من التباينات $\text{ع}^{٢}_{\text{ى}}$. ففى تحديد حجوم العينات نستخدم نفس الصيغة (١١) وهى

$$\text{ن}_{\text{ى}} = \frac{\text{ن}}{\text{هـ}} \times \text{هـ}_{\text{ى}} \qquad (٢٠)$$

فى حالة استخدام طريقة التقسيم المتناسب . أما فى حالة التقسيم الأمثل فنستخدم الصيغة (١٢) بعد وضعها كالآتى :

$$\text{ن}_{\text{ى}} = \frac{\text{ن}}{\text{مج}\ \text{هـ}_{\text{ى}} \sqrt{\text{ح}_{\text{ى}} (١ - \text{ح}_{\text{ى}})}} \times \text{هـ}_{\text{ى}} \sqrt{\text{ح}_{\text{ى}} (١ - \text{ح}_{\text{ى}})} \qquad (٢١)$$

حيث $\text{ح}_{\text{ى}}$ هو نسبة وقوع الحدث فى الطبقة ى .

وفى تقدير النسبة ح وهى احتمال وقوع الحدث فى المجتمع نستخدم الصيغة (١٣) بعد وضعها فى الصورة الآتية :

$$ر = \frac{١}{هـ} \sum_{١} هـ_ى ر_ى \qquad (٢٢)$$

وحين تكون المعاينة بطريقة التقسيم المتناسب تؤول هذه الصيغة إلى :

$$ر = \frac{١}{ن} \sum_{١} ن_ى ر_ى \qquad (٢٣)$$

وهذا يعنى أننا فى هذه الحالة نقدر الاحتمال ح فى المجتمع بواسطة النسبة المشاهدة فى العينة الكلية التى تتألف من ضم جميع العينات الجزئية .

كذلك ، لتقدير الخطأ المعيارى $\sigma_ر$ لتوزيع المعاينة للنسبة ر نستخدم الصيغة (١٦) بعد وضعها كالآتى :

$$ع_خ = \sqrt{\sum_{١} و_ى^٢ \times \frac{ر_ى (١ - ر_ى)}{ن_ى} (١ - \frac{ن_ى}{هـ_ى})} \qquad (٢٤)$$

حيث $و_ى = هـ_ى / هـ$

وحين تكون المعاينة بطريقة التقسيم المتناسب وعلى فرض تساوى تباينات الطبقات نستخدم الصيغة (١٨) وهى

$$ع_خ = \frac{ع_د}{\sqrt{ن}} \sqrt{١ - \frac{ن}{هـ}} \qquad (٢٥)$$

$$\text{حيث } ع_د^٢ = \frac{١}{ن - ك} \sum (ن_ى - ١) ر_ى (١ - ر_ى) \qquad (٢٦)$$

هو تقدير للتباين المشترك للطبقات .

مثال (١٥ - ٧) :

كان عدد الأطفال فى إحدى المدن الكبرى ٤٣١٥٤٢ طفلا . وفى إحدى التجارب كان المطلوب تحديد فترة ثقة بدرجة ٩٥٪ لنسبة الإصابة بمرض ما بين هؤلاء الأطفال عن طريق عينة من ٤٥٠ طفلا أى بواقع $\frac{٤٥٠}{٤٣١٥٤٢} = ٠,٠٠١٤$ تقريبا . وقد قسم المجتمع إلى ثلاث طبقات بحسب كثافة السكان ، إذ كان من المعتقد أن الإصابة بهذا المرض تختلف باختلاف هذه الكثافة . وقد وجد أن أعداد الأطفال فى هذه الطبقات ٢٢٠٣٨٦ ، ١٤٦٠٣١ ، ٦٥١٢٥ .

الحل :

نظرا لعدم وجود معلومات عن تباينات الإصابة بالمرض فى الطبقات فقد استخدم لتحديد أحجام العينات طريقة التقسيم المتناسب بحسب الصيغة (١١) وهى

$$ن_ى = \frac{ن}{هـ} \times هـ_ى .$$

$$\therefore ن_١ = \frac{٤٥٠}{٤٣١٥٤٢} \times ٢٢٠٣٨٦ = ٢٣٠ \text{ طفلا}$$

$$، ن_٢ = \frac{٤٥٠}{٤٣١٥٤٢} \times ١٤٦٠٣١ = ١٥٢ \text{ طفلا}$$

$$، ن_٣ = \frac{٤٥٠}{٤٣١٥٤٢} \times ٦٥١٢٥ = ٦٨ \text{ طفلا}$$

وقد اختيرت عينات عشوائية بسيطة من تلك الطبقات بحسب الأحجام الناتجة . وبفحص الأطفال وجد أن أعداد الأطفال المصابين بالمرض ونسبة هذه الإصابة فى العينات كما هو مبين بالعمودين الآخرين من الجدول (١٥ - ٥) الآتى :

الجدول (١٥ – ٥)

أعداد المصابين بالمرض ونسب هذه الإصابة

نوع الطبقة	حجم الطبقة $هـ_ى$	حجم العينة $ن_ى$	عدد المصابين فى العينة	نسبة الإصابة٪ $س_ى$
شديدة الازدحام	٢٢٠٣٨٦	٢٣٠	١٦٤	٧١,٣
متوسطة الازدحام	١٤٦٠٣١	١٥٢	٩٢	٦٠,٥
قليلة الازدحام	٦٥١٢٥	٦٨	١٨	٢٦,٥
المجموع	٤٣١٥٣٢	٤٥٠	٢٧٤	٦٠,٩

نظرا لأننا استخدمنا طريقة التقسيم المتناسب فإن نسبة الإصابة بالمرض فى مجتمع الأطفال تقدر بنسبة الإصابة فى العينة الكلية وهى

$$س = ٦٠,٩٪$$

نظرا لافتراضنا أن التباينات متساوية فى الطبقات فإننا نقدر التباين المشترك بالصيغة (٢٦) كما يلى :

$$ع_{د}^{٢} = \frac{١}{٤٤٧} (٢٢٩ \times ٠,٧١٣ \times ٠,٢٨٧ + ١٥١ \times ٠,٦٠٥ \times ٠,٣٩٥$$

$$+ ٦٧ \times ٠,٢٦٥ \times ٠,٧٣٥)$$

$$= ٠,٢١٤٨$$

∴ $ع_ر = ٠,٤٦٣$

∴ الخطأ المعيارى لتوزيع المعاينة للنسبة ر للعينات التى من الحجم ٤٥٠ هو حسب الصيغة (٢٥) :

$ع_غ = \frac{٠,٤٦٣}{\sqrt{٤٥٠}} = ٠,٠٢٢$ (مع إهمال عامل التصحيح لقربه من الواحد)

فى عينة بالحجم ٤٥٠ يكون المتوسط النسبى موزعا توزيعا معتدلا على وجه التقريب وبالتالى يكون حدا الثقة بدرجة ٩٥٪ لمتوسط الاصابة بالمرض فى المجتمع هما ر ± $ع_غ$ × ١,٩٦ أى ٠,٦٠٩ ± ٠,٠٢٢ × ١,٩٦ . وبحساب هاتين القيمتين نجد أن الفترة المطلوبة هى (٥٦,٦٪ ، ٦٥,٢٪) .

(١٥ – ٥) العينة المتعددة المراحل MULTISTAGE SAMPLE

(أو العينة العشية nested sample)

حين يكون المجتمع كبيرا نضطر أحيانا إلى اختيار العينة عن طريق سلسلة من المراحل . وكمثال لذلك نفرض أننا نريد اختيار عينة لتقدير عدد الحالات من المرضى الذين فحصوا بالأشعة فى أسبوع فى المستشفيات الحكومية بدولة ما . فى هذه الحال يصعب بل يستحيل تصميم خطة للمعاينة من المرضى مباشرة ، ولذلك نلجأ إلى المعاينة على مراحل كما يلى . نجرى حصرا بالمحافظات أو المناطق الجغرافية التى بها مستشفيات حكومية . تبدأ المعاينة باختيار عينة عشوائية بسيطة من هذه المناطق حجمها $ن_١$ منطقة ونسجل أسماء المستشفيات الحكومية بكل منها ، وهذه هى المرحلة الأولى . نأخذ من كل منطقة من المناطق السابق اختيارها عينة عشوائية بسيطة من المستشفيات حجمها $ن_٢$ من المستشفيات وهذه هى المرحلة الثانية وبذلك يكون لدينا $ن_١ \times ن_٢$ مستشفى . نأخذ من كل من هذه المستشفيات

عينة عشوائية بسيطة من المرضى الذين دخلوها أو كانوا مقيمين بها فى الأسبوع المحدد وليكن حجمها $ن_{٣}$ مريضا وهذه هى المرحلة الثالثة والأخيرة ، وبذلك نكون قد حصلنا على عينة إجمالية حجمها $ن_{١} \times ن_{٢} \times ن_{٣}$ من المرضى ، ونستطيع حينئذ أن نفحص ملفاتهم لمعرفة عدد الذين فحصوا بالأشعة .

وقد يكون من المناسب أحيانا استخدام التقسيم الطبقى فى واحدة أو أكثر من مراحل المعاينة إذا استدعى الأمر ذلك فتقسم المناطق الجغرافية مثلا إلى مدن كبيرة ومدن صغيرة وقرى ، أو تقسم المستشفيات بحسب التخصص ، أو يقسم المرضى بحسب الجنس .

وكمثال آخر ، نفرض أننا نريد تقدير متوسط طول فتلة القطن فى بالة كبيرة من القطن . نأخذ عدة حفنات من القطن عشوائيا من جوانب مختلفة من البالة وهذه مرحلة أولى . نأخذ كل حفنة من الحفنات التى اخترناها ونقسمها إلى جزءين نرمى أحدهما ونحتفظ بالآخر وهذه مرحلة ثانية . نكرر هذه العملية عدة مرات حتى نحصل على عدد مناسب من الفتلات لقياسها وحساب متوسط الطول فيها .

يلاحظ فى هذا المثال أن الوحدات المختارة فى كل مرحلة على هيئة « مجموعات » aggregates وليست على هيئة مفردات . فى مثل هذه الحالة توصف المعاينة بأنها معاينة **عنقودية** cluster sampling ومن أمثلتها أيضا سحب عينات من رمال أحد الشواطىء أو سحب عبوات من ماء مجرى نهر ، أو سحب فصول كاملة من عدد من المدارس .

ومن الحالات التى تستلزم المعاينة المتعددة المراحل تلك التى تحتاج إلى إجراء الاختبارات الكيميائية أو الفيزيائية أو البيولوجية التى يصعب إجراؤها إلا على أجزاء صغيرة من المادة المختبرة كما فى المثال (١٥ — ٨) الآتى .

التحليل الإحصائى :

يحتاج تحليل العينات متعددة المراحل إلى استخدام أسلوب تحليل التباين بالنموذج عشوائى التأثيرات – راجع البند (٨ – ١٤) – ولنرى ذلك نبدأ بتناول العينة ذات المرحلتين مستعينين بالمثال (٨ – ١٦) حيث كان اهتمامنا بتقدير نسبة الكلسيوم في أوراق اللفت الأخضر وكانت المعاينة على مرحلتين أولهما أخذ عينة عشوائية من أوراق النبات ، وتسمى وحدات هذه العينة بالوحدات الابتدائية primary sampling units ، وثانيهما أخذ عينة عشوائية من أجزاء كل ورقة وقياس نسبة الكلسيوم ، وتسمى وحدات هذه العينة بوحدات المرحلة الثانية second-stage units, or sub-units . ولقد اعتبرنا أن الوحدات الابتدائية هي عينة عشوائية من « المعالجات » وأن نسب الكلسيوم هى القيم الناتجة تحت تأثير هذه المعالجات . ومن ثم كان تحليلنا للبيانات عن طريق تحليل التباين بالنموذج عشوائى التأثيرات الذى يأخذ الصيغة الآتية :

$$س_{ر ى} = \mu + ا_{ى} + خ_{ر ى} \qquad (٢٧)$$

لقد قدمنا فى البند (٨ – ١٤) تفصيلا لهذا التحليل مدعما بالمثالين (٨ – ١٦) و(٨ – ١٧) وليس هناك ما يدعو لتكرار ذلك هنا .

نقوم الآن بتحليل عينة ذات ثلاث مراحل ونستعين فى ذلك بالمثال (١٥ – ٨) الآتى .

مثال (١٥ – ٨) :

اعتبر تجربة المثال (٨ – ١٦) عن محتوى الكلسيوم فى أوراق نبات اللفت الأخضر . وافرض أننا لم نبدأ باختيار عينة عشوائية من الأوراق بل بدأنا بعينة عشوائية من نبات اللفت ذاته حجمها ك =٤. نباتات (الوحدات الابتدائية) ثم أخذنا من كل نبات عينة عشوائية من ا = ٣ ورقات (وحدات المرحلة الثانية)

ثم أخذنا من كل ورقة عينة عشوائية من ب = ٢ من الأجزاء وزن كل منها ١٠٠ ملليجرام (وحدات المرحلة الثالثة) فحصلنا بذلك على ٤ × ٣ × ٢ = ٢٤ جزءا من أوراق النبات هي التي نقوم بقياس محتوى الكلسيوم فيها . نفرض أن القياسات جاءت كما فى الجدول (١٥ - ٦) الآتى .

الجدول (١٥ - ٦)

النسب المئوية للكالسيوم فى ب = ٢ جزءا من كل من ا = ٣ ورقة من كل من ك = ٤ نباتا

النبات	(١)			(٢)			(٣)			(٤)			
الأوراق	ا	ب	حـ	ا	ب	حـ	ا	ب	حـ	ا	ب	حـ	
نسبة الكالسيوم فى جزئى كل ورقة	٣,٢٨ ٣,٠٩	٣,٥٢ ٣,٤٨	٢,٨٨ ٢,٨٠	٢,٤٦ ٢,٤٤	١,٨٧ ١,٩٢	٢,١٩ ٢,١٩	٢,٧٧ ٢,٦٦	٣,٧٤ ٣,٤٤	٢,٥٥ ٢,٥٥	٣,٧٨ ٣,٨٧	٤,٠٧ ٤,١٢	٣,٣١ ٣,٣١	
المجموع للورقة	٦,٣٧	٧,٠٠	٥,٦٨	٤,٩٠	٣,٧٩	٤,٣٨	٥,٤٣	٧,١٨	٥,١٠	٧,٦٥	٨,١٩	٦,٦٢	
المجموع للنبات	١٩,٠٥			١٣,٠٧			١٧,٧١			٢٢,٤٦			٧٢,٢٩

(فى هذا المثال أخذنا عددا متساويا من الأوراق من كل نبات وكان من الممكن أى يختلف هذا العدد من نبات إلى آخر . وكذلك بالنسبة لعدد الأجزاء التى أخذت من كل ورقة .)

إذا كانت $س_{ر ى ه}$ ترمز إلى نسبة الكلسيوم فى الجزء ر من الورقة ى من النبات هـ فإن النموذج الإحصائى للتحليل هو امتداد للنموذج (٢٣) :

$$س_{ر ى ه} = \mu + أ_{ه} + ب_{ى ه} + خ_{ر ى ه} \qquad (٢٨)$$

(ر = ١ ، ٢ و ى = ا ، ب ، حـ و هـ = ١ ، ٢ ، ٣ ، ٤

حيث $أ_ى$ تشير إلى النباتات ، $ب_{ى و}$ تشير إلى الأوراق .
ولإمكانية التحليل الإحصائى سنفترض كالمعتاد أن

$أ_ى$: مع (٠ ، σأ) ، $ب_{ى و}$: مع (٠ ، σب) ، $خ_{ى و هـ}$: مع (٠ ، σ) (٢٩)

فى تحليل التباين نفصل مجموع المربعات الكلى للقياسات (نسب محتوى الكلسيوم) إلى مصادر مستقلة للاختلاف . وفى هذا المثال نجد أن هذه المصادر هى : النباتات – أوراق نفس النبات – محتوى الكلسيوم بأجزاء نفس الورقة . ولايجاد الاختلافات فى هذه المصادر نحسب بالطريقة المعتادة كلا من م م (الكلى) ، م م (بين النباتات) ، م م (بين الأوراق) أما الاختلافان الباقيان فنحسبهما من هذه الاختلافات كالآتى :

(١) م م (بين أوراق نفس النبات) = م م (بين الأوراق) – م م (بين النباتات)

(٢) م م (بين القياسات على نفس الورقة) = م م (الكلى) – م م (بين الأوراق)

وفى المثال نجد من الجدول (١٥ – ٦) ما يلى .

عامل التصحيح = $\frac{م^2}{ن}$ = $\frac{٧٢,٢٩^2}{٢٤}$ = ٢١٧,٧٤٣٥

م م (الكلى) = $٣,٢٨^2 + ٣,٠٩^2 + \ldots + ٣,٣١^2 + ٣,٣١^2$ ٢١٧,٧٤٣٥

= ١٠,٢٧٠٤ بدرجات حرية ن – ١ = ٢٣

م م (بين النباتات) = $\frac{1}{٦}$ ($١٩,٠٥^2 + ١٣,٠٧^2 + ١٧,٧١^2 + ٢٢,٤٦^2$)
– ٢١٧,٧٤٣٥

= ٧,٥٦٠٣ بدرجات حرية ك – ١ = ٣

م م (بين الأوراق) = $\frac{1}{2}$(٦,٣٧² + ٧,٠٠² + ... + ٨,١٩² + ٦,٦٢²)
–: ٢١٧,٧٤٣٥

= ١٠,١٩٠٥ بدرجات حرية أ ك − ١ = ١١

∴ م م (بين أوراق نفس النبات) = ١٠,١٩٠٥ − ٧,٥٦٠٣

= ٢,٦٣٠٢ بدرجات حرية أ ك − ك = ٨

، م م (بين القياسات على نفس الورقة) = ١٠,٢٧٠٤ − ١٠,١٩٠٥

= ٠,٠٧٩٩ بدرجات حرية ن − أ ك = ١٢

وينتج جدول التباين الآتى :

الجدول (١٥ − ٧)

مصدر التباين	م م	د ح	تقدير التباين	التباين المتوقع
بين النباتات	٧,٥٦٠٣	٣	ع² = ٢,٥٢٠١	σ^2 + ب $\sigma_ب^2$ + أ ب $\sigma_أ^2$
بين الأوراق داخل النباتات	٢,٦٣٠٢	٨	ع²ب = ٠,٣٢٨٨	σ^2 + ب $\sigma_ب^2$
بين القياسات داخل الأوراق داخل النباتات	٠,٠٧٩٩	١٢	ع² = ٠,٠٠٦٧	
الكل	١٠,٢٧٠٤	٢٣		

من العمود الأخير نلاحظ أن كل مركبة من مركبات التباين داخلة فى المركبة السابقة لها ، ولهذا يمكن بسهولة أن نرى ما يلى :

$\text{ع}^2_{\text{خ}}$ هو تقدير غير متحيز للتباين σ^2 (٣٠)

$\frac{1}{\text{ب}}(\text{ع}^2_{\text{ب}} - \text{ع}^2_{\text{خ}})$ هو تقدير غير متحيز للتباين $\sigma^2_{\text{ب}}$ (٣١)

$\frac{1}{\text{ا ب}}(\text{ع}^2_{\text{ا}} - \text{ع}^2_{\text{ب}})$ هو تقدير غير متحيز للتباين $\sigma^2_{\text{ا}}$ (٣٢)

ففى هذا المثال نجد التقديرات الآتية :

(١) تقدير σ^2 = ٠,٠٠٦٧

(٢) تقدير $\sigma^2_{\text{ب}}$ = $\frac{1}{2}$ (٠,٣٢٨٨ − ٠,٠٠٦٧) = ٠,١٦١١

(٣) تقدير $\sigma^2_{\text{ا}}$ = $\frac{1}{6}$ (٢,٥٢٠١ − ٠,٣٢٨٨) = ٠,٣٦٥٢

كما نجد التقديرين الآتيين :

(٤) يقدر الوسط الحسابى μ للنسبة المئوية للكلسيوم فى المجتمع بالوسط الحسابى للعينة وهو :

$$\bar{\text{س}} = \frac{٧٢,٢٩}{٢٤} = ٣,٠١$$

(٥) يقدر الخطأ المعيارى لتوزيع المعاينة للوسط الحسابى كالآتى :

$$\text{تقدير } \sigma^2_{\bar{\text{س}}} = \frac{\text{ع}^2_{\text{ا}}}{\text{ن}} = \frac{\text{التباين بين النباتات (الوحدات الابتدائية)}}{\text{حجم العينة}} \quad (٣٣)$$

= ٢,٥٢٠١ ÷ ٢٤ = ٠,١٠٥

ومن الجذر التربيعى لهذه القيمة نستطيع حساب فترات الثقة للوسط الحسابى لنسبة محتوى الكلسيوم فى المجتمع .

الاختبارات الإحصائية :

يهمنا هنا إجراء الاختبارين الآتيين :

(أولا) اختبار ما إذا كان محتوى الكلسيوم يختلف من نبات إلى آخر .

وهذا يعنى اختبار الفرض الصفرى $\sigma^{2}_{ا} = ٠$ – راجع البند (٨ – ١٤) . ومن الجدول (١٥ – ٧) نرى أنه إذا كان هذا الفرض صحيحا فإن $ع^{2}_{ا}$ ، $ع^{2}_{ب}$ يكونان تقديرين مستقلين لنفس التباين وإذن الاختبار المناسب لهذا الفرض هو اختبار ف بالصورة الآتية :

$$ف = \frac{ع^{2}_{ا}}{ع^{2}_{ب}} \quad \text{بدرجتى حرية ك – ١ ، ا ك – ك} \qquad (٣٤)$$

$$\therefore \quad ف_{ى} = \frac{٢,٥٢٠١}{٠,٣٢٨٨} = ٧,٦٦^{**} \qquad \text{بدرجتى حرية ٣ ، ٨}$$

وهذه القيمة تزيد عن القيمة الحرجة $ف_{٠,٠١[٣ ، ٨]} = ٧,٥٩$ مما يدعونا إلى رفض الفرض الصفرى عند المستوى ٠,٠١ واستنتاج أن محتوى الكلسيوم لا يتساوى فى جميع النباتات .

(ثانيا) اختبار ما إذا كان محتوى الكلسيوم يختلف من ورقة إلى أخرى . وهذا يعنى اختبار الفرض الصفرى $\sigma^{2}_{ب} = ٠$ ومن الجدول (١٥ – ٧) نرى أنه إذا كان هذا الفرض صحيحا فإن $ع^{2}_{ب}$ ، $ع^{2}_{خ}$ يكونان تقديرين مستقلين للتباين σ^{2} وإذن الاختبار المناسب هو :

$$ف = \frac{ع^{2}_{ب}}{ع^{2}_{خ}} \quad \text{بدرجتى حرية ا ك – ك ، ن – ا ك} \qquad (٣٥)$$

∴ $ف_ى = \frac{٠,٣٢٨٨}{٠,٠٠٦٧} = ٤٩^{**}$ بدرجتى حرية ٨ ، ١٢

وهذه القيمة تزيد كثيرا عن القيمة الحرجة $ف_{٠,٠١ [٨ ، ١٢]}$ = ٤,٥٠ مما يجعلنا نرفض الفرض الصفرى عند مستوى عالى من الدلالة ونحكم بأن محتوى الكلسيوم لا يتساوى فى جميع الأوراق .

(١٥ – ٦) العينة المنتظمة SYSTEMATIC SAMPLE

تتخذ المعاينة المنتظمة الأسلوب المبين بالمثال الآتى . نفرض أن أحد الجيولوجيين مهتم بدراسة محتوى المعادن الثقيلة فى البطانة المعدنية mineral suit لمنطقة رملية بها ه = ١٠٠٠ طبقة بارزة من الصخور outrcops ممتدة فى صف على سطح الارض ، والمطلوب اختيار عينة حجمها ن = ٥٠ صخرة لتحليلها وفصل المعادن منها ، على أن تؤخذ وحداتها بحيث تكون موزعة توزيعا متعادلا على المنطقة . لتحقيق ذلك نرقم الصخور من ١ إلى ١٠٠٠ ثم نقسمها إلى ن = ٥٠ قسما بحيث يحتوى كل قسم على نفس العدد من الصخور أى على ك = ٢٠ صخرة (١٠٠٠ ÷ ٥٠ = ٢٠) . نختار عشوائيا عددا يقع بين ١ ، ٢٠ ليحدد رقم الصخرة التى تؤخذ من القسم الأول . فمثلا إذا كان العدد العشوائى ٣ يكون هذا العدد هو رقم الصخرة الأولى فى العينة ، ثم نأخذ بانتظام أعدادا يزيد كل منها عن سابقه بمقدار ٢٠ فتكون أرقام الصخور التى تدخل فى العينة هى ٣ ، ٢٣ ، ٤٣ ، ٦٣ ، ٨٣ ، ١٠٣ ، ... ، ٩٨٣ . أما إذا كان العدد العشوائى الذى اختير فى البداية هو ١٥ فإن أرقام الصخور تكون فى هذه الحال ١٥ ، ٣٥ ، ٥٥ ، ٧٥ ، ٩٥ ، ١١٥ ، ... ، ٩٩٥ .

فى هذا المثال كان حجم المجتمع ه مضاعفا صحيحا لعدد الأقسام (ه = ن ك) ولذلك فإن أى عينة تختار بهذه الطريقة تأخذ نفس الحجم ن . أما اذا

كان هـ $\neq$ ن ك فإن العينات لا تكون جميعها من حجم واحد بل قد يزيد حجم بعضها بواحد عن البعض الآخر . فمثلا نفرض أن هـ = ٢٣ ، ك = ٥ . إن العينات الخمسة التى يمكن اختيارها تكون أرقام وحداتها كما فى الجدول (١٥ - ٨) الآتى :

الجدول (١٥ - ٨)

(١)	(٢)	(٣)	(٤)	(٥)
١	٢	٣	٤	٥
٦	٧	٨	٩	١٠
١١	١٢	١٣	١٤	١٥
١٦	١٧	١٨	١٩	٢٠
٢١	٢٢	٢٣		

حيث نلاحظ أن حجم كل من العينات الثلاث الأولى ن = ٥ بينما حجم كل من العينتين الأخيرتين ن = ٤ . وهذه الحقيقة تسبب إزعاجاً في تحليل بيانات العينة المنتظمة .

إذا كان هـ = ن ك (حيث تتساوى حجوم العينات التى تؤخذ) يكون الوسط الحسابى للعينة تقديراً غير متحيز للوسط الحسابى للمجتمع . أما إذا كان هـ $\neq$ ن ك فإن هذا التقدير يكون متحيزاً ، غير أنه يمكن إزالة هذا التحيز بإعطاء احتمالا أكبر لاختيار بعض العينات . ففى الجدول (١٥-٨) إذا أعطينا الاحتمال $\frac{٥}{٢٣}$ لاختيار كل من العينات الثلاث الأولى والاحتمال $\frac{٤}{٢٣}$ لاختيار كل من العينتين الأخيرتين فإن الوسط الحسابى للعينة يكون حينئذ تقديراً غير متحيز لمتوسط المجتمع . (لاحظ أن $\frac{٥}{٢٣} \times ٣ + \frac{٤}{٢٣} \times ٢ = \frac{١٥+٨}{٢٣} = ١٫٠$)

أما تقدير تباين المجتمع أو تقدير الخطأ المعيارى لتوزيع المعاينة للوسط الحسابى فلا توجد طريقة موثوق بها لايجادهما من بيانات مشاهدة فى عينة وهذا هو العيب الرئيسي فى المعاينة المنتظمة .

أما العيب الثانى فإن العينة المنتظمة تكون متحيزة ولا تعبر تعبيرا صادقا عن المجتمع إذا كان هناك نوع من الاختلافات الدورية أو الموسمية فى وحدات المجتمع خاصة إذا حدث أن كانت الوحدات المختارة قريبة من مراكز هذه الاختلافات .

ومن الواضح أن العينة العشوائية المنتظمة أسهل وأسرع فى اختيارها من أى عينة أخرى وأقل تعرضا للخطأ إذ يكفى تحديد عدد عشوائى واحد . ولذلك فهى تستخدم حين تكون أقل تكلفة بكثير من أى طريقة أخرى للمعاينة ، أو حين يكون المطلوب تغطية المجتمع بشكل متعادل فهى فى هذه الحالة تعطى نتائج أكثر دقة من العينة العشوائية البسيطة .

هذا مع ملاحظة أنه فى المعاينة المنتظمة يكون لكل وحدة من وحدات المجتمع نفس الفرصة فى الدخول فى العينة ، وهى فى هذه الصفة تشبه المعاينة العشوائية البسيطة . غير أن احتمال الحصول على عينة منتظمة من حجم ما لا يكون مساويا لاحتمال الحصول على عينة منتظمة أخرى من نفس الحجم (اختيرت بتغيير العدد العشوائى الابتدائى) كما هو الحال فى المعاينة العشوائية البسيطة ، وهذا فرق كبير بين هذين النوعين من المعاينة . والواقع أن العينة المنتظمة ليست عشوائية إلا فى اختيار العدد الابتدائى مما يعوق عملية التحليل الاحصائى .

(١٥ – ٧) المعاينة المساحية Area Sampling

المعاينة المساحية هى تلك التى تختار فيها العينات من مسطح من الأرض ، وهى تستخدم فى كثير من الدراسات السكانية وفى علوم الزراعة والجيولوجيا وغيرها .

وطريقة المعاينة المساحية ليس بها جديد من حيث المبدأ إذ تؤخذ العينات بنفس الطرق سابقة الذكر بحسب طبيعة الدراسة التى تجرى ، فقد تكون عشوائية بسيطة أو طبقية أو على مراحل أو منتظمة . ولا يحتاج الأمر إلا إلى إدخال تعديل فى خطة المعاينة يمكننا من تحديد الوحدات التى تدخل فى العينة .

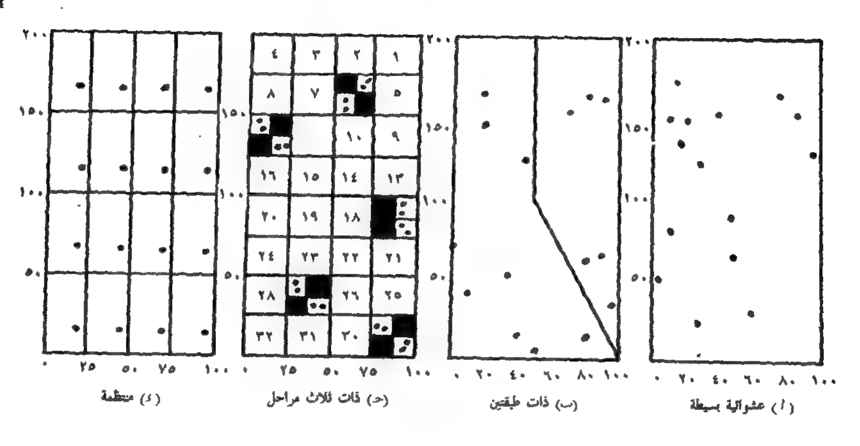
ومن الإجراءات المفيدة هنا رسم خريطة مصغرة للمسطح المعطى على مستوى يحدد بمحورين متعامدين أحدهما يعين مثلا الشمال والجنوب والآخر يعين الشرق والغرب مع تحديد نقطة أصل مناسبة . وعند المعاينة تختار النقط (أو القطع أو المربعات) التى تدخل فى العينة عن طريق الاختيار العشوائى لأزواج من الأعداد تتخذ كإحداثيات للنقط التى تحدد على الخريطة ومن ثم على المسطح الأصلى . والمعتاد اختيار وحدات هذه الأزواج من الأعداد عن طريق جداول الأرقام العشوائية .

مثال (١٥ – ٩) :

لدينا منطقة من الأرض مستطيلة الشكل عرضها ١٠٠ مترا وطولها ٢٠٠ مترا ونريد دراسة بعض الخواص الكيميائية لتربة هذه الأرض عن طريق اختيار عينة منها ، فى الحالات الأربع الآتية :

(أولاً) : اختيار عينة عشوائية بسيطة من ١٥ نقطة .

نستخدم جداول الأرقام العشوائية لاختيار ١٥ عددا عشوائيا يقع بين ٠ ، ١٠٠ مثلا ٧٤ ، ١٠ ، ١١ ، ٤٤ ، ١٦ ، ... ثم لاختيار ١٥ عددا عشوائيا يقع بين ٠ ، ٢٠٠ مثلا ١٦٨ ، ١٥٠ ، ١٧٤ ، ٩١ ، ١٣٦ ، ... وبهذا يتحدد لنا ١٥ زوجا من الأعداد هى (٧٤ ، ١٦٨) ، (١٠ ، ١٥٠) ، (١١ ، ١٧٤) ، (٤٤ ، ٩١) ، (١٦ ، ١٣٦) ، ... كل منها يحدد نقطة فى المستوى ، فتكون النقطة الناتجة هى وحدات العينة المطلوبة . انظر الشكل (١٥ –١ –ا) .



الشكل (١٥ - ١) : عينات عشوائية مساحية

(ثانيا) : اختيار عينة طبقية من ١٥ نقطة إذا كان من المعروف أن الأرض مقسمة إلى طبقتين مختلفتين النسبة بينهما ٢ : ٣ على وجه التقريب .

نسحب من الطبقتين عينتين عشوائيتين بسيطتين يتناسب حجماهما مع حجمى الطبقتين ، أى نسحب $\frac{٢}{٥}$ × ١٥ = ٦ وحدات من الطبقة الأولى ، $\frac{٣}{٥}$ × ١٥ = ٩ وحدات من الطبقة الثانية . وعلى ذلك نختار ٦ أزواج من الأعداد العشوائية بحيث تقع النقط الممثلة لها فى الطبقة الأولى ، ونختار ٩ أزواج من الأعداد العشوائية بحيث تقع النقط الممثلة لها فى الطبقة الثانية فنحصل مثلا على النقط الآتية :

للطبقة الأولى : (٩١ ، ١٦٣) ، (٩٥ ، ٣٥) ، (٨١ ، ٦١) ، (٨٠ ، ١٦٣) ، (٩٠ ، ٦٧) ، (٧٠ ، ١٥٥) .

للطبقة الثانية : (١٩ ، ١٦٦) ، (٤١ ، ١٦) ، (٠ ، ٦٩) ، (٣٢ ، ٥٣) ، (١٩ ، ١٤٧) ، (٥٢ ، ٥) ، (١١ ، ٤٠) ، (٤٥ ، ١٢٧) ، (٨٠ ، ١٥) .

انظر الشكل (١٥ — ١ — ب) .

(ثالثا) : اختيار عينة عشوائية من ٢٠ نقطة تؤخذ على ثلاث مراحل .

نقسم الأرض إلى عدد من المربعات المتساوية المساحة . فمثلا إذا أخذنا طول المربع ٢٥ فإن الأرض تنقسم إلى ٤ × ٨ = ٣٢ مربعا . نرقم هذه المربعات من ١ إلى ٣٢ . نبدأ بأخذ عينة عشوائية من ٥ مربعات ، مثلا المربعات ذوات الأرقام ١٧ ، ٢٧ ، ٦ ، ٢٩ ، ١٢ وهذه هى المرحلة الأولى (الوحدات الابتدائية) . نقسم كلا من هذه المربعات الخمسة إلى ٤ أجزاء متساوية المساحة ونأخذ من كل مربع جزءين عشوائيا ، وهذه هى المرحلة الثانية وتحتوى على ٥ × ٢ = ١٠ أجزاء كل منها ربع مربع . وأخيرا نختار من كل ربع مربع نقطتين عشوائيا فنحصل على ٢٠ نقطة هى التى تمثل العينة المطلوبة — انظر الشكل (١٥ — ١ — حـ) .

(رابعا) : اختيار عينة منتظمة من ١٦ نقطة .

نقسم كلا من الطول والعرض إلى ٤ أقسام متساوية الطول فنحصل على ١٦ قسما كل منها مستطيل عرضه ٢٥ مترا وطوله ٥٠ مترا . نحدد عددا عشوائيا بين ٠ ، ٢٥ وليكن ٢٢ ونحدد عددا عشوائيا بين ٠ ، ٥٠ وليكن ١٨ فيكون إحداثيا النقطة الابتدائية (٢٢ ، ١٨) . تحدد النقط الأخرى بانتظام بحيث تبعد كل نقطة عن سابقتها بمسافة قدرها ٢٥ على المحور الأفقى ، ٥٠ على المحور الرأسى . انظر الشكل (١٥ — ١ — ى) .

(١٥ — ٨) العينات غير الاحتمالية :

نعلم أن العينة غير الاحتمالية هى تلك التى نأخذها من المجتمع دون أن نعرف احتمالات دخول وحدات المجتمع فيها ومن ثم لا نستطيع إخضاعها لقواعد الاحتمالات ولا أن نطبق عليها الاختبارات الإحصائية سالفة الذكر . ومع ذلك لا يجوز التقليل من أهمية هذه العينات فكثير منها له استخدامات هامة ويمكن أن تعطى مؤشرات مفيدة عن المجتمعات التى تؤخذ منها .

ومن هذه العينات مايسمى **بالعينة الغرضية** Purposeful sample وهى تلك العينة غير العشوائية التى لا تختار بهدف التحليل الإحصائى المعتاد بل لأداء مهمة أو غرض محدد كما هو الحال فى البحوث الاستطلاعية لتقدير تكاليف البحث أو تلمس المشكلات المتوقعة أو لتدريب المساعدين على عملية جمع البيانات . وهنا يختار الباحث الجزء من المجتمع القريب من متناول يده دون تحمل مشقة المعاينة العشوائية .

وهناك ما يسمى **بالعينة بالحصص** Quota sample وهذا النوع تستخدمه كثير من المؤسسات الصحفية ومعاهد استطلاع الرأى ، ومن أشهرها معهد جالوب Gallup بالولايات المتحدة الأمريكية الذى يستشف نتائج الانتخابات العامة قبل

إجرائها بسرعة وتكاليف قليلة ، فيطلب من عدد من العاملين استطلاع رأى عدد معين من الناس (حصة) فى أحد الأحياء أو المناطق فيقوم كل عامل بسؤال من يصادفه من الناس فى المكان المحدد له حتى يتم الحصة المنوطة به .

كما أن هناك **عينات اضطرارية** كما هو الحال فى عينة تتألف من متطوعين فى الدراسات التى تكون فيها القياسات أو التجارب متعبة أو غير مستحبة أو تحتمل الضرر للأفراد الذين تجرى عليهم الدراسة .

وهناك مايسمى **بعينة التفتيش** Search sample وهى تلك التى تهدف إلى التفتيش عن معلومات جديدة كتصيد أنواع جديدة من الحشرات أو القواقع أو الصخور المعدنية ، أو الكشف عن رواسب جيرية تصلح لصناعة الأسمنت ، أو التنقيب عن الآبار والمياه الجوفية مما يفتح آفاقا جديدة للدراسة النظرية والتطبيقات العملية .

ملحق (١)
أجوبة التمارين

تمارين (١) :

(٢) ٦٥,٠   ٠,٠٦٤   ١٤٤,٠   ٣,٠   ٣,٤٥   ٢٥,٠

(٣) ٤٩,٤٨   ٤٩,٥

(٤) $\frac{١}{٣}$ $\frac{٢}{١٥}$ $\frac{٣}{١٥}$

(٥) $\frac{١٩}{٢٥}$

تمارين (٢ – ١) :

(١) ١٠٩   ١٣,١٥٧   ١٢,٠٧

(٢) ٤٨,٨٩   ١٠,٣٢   ٢١,١١

(٣) ٣,٠٢٥   ١,٧٦٥

(٤) ٤ م م × $١٠^{-١}$ ، ٠,٣٤٦ م م × $١٠^{-١}$ ، ٦٢,٦٪ ، ٤

(٦) $ر_{١}$ = ٢,٢٧٥ ، $ر_{٢}$ = ٢,٥ ، $ر_{٣}$ = ٢,٩٨

تمارين (٢ – ٢) :

(١) التوزيع ملتوى إلى اليمين .

(٢) فى توزيع غير المدخنين يتجمع عدد كبير فى وسط التوزيع (بين ١٩،٢١) ويتناقص هذا العدد تدريجيا عند الطرفين ، وبالعكس فى توزيع المدخنين .

(٣) نعم الفيران تتعلم من التدريب — فى الفيران المدربة يتجمع عدد كبير من الفروق حول القيمة ٤ وهى فى وسط التوزيع ويتناقص العدد تدريجيا فى الطرفين ، وبالعكس فى الفيران غير المدربة .

تمارين (٣ - ١) :

(١) ٩,٧٧ × ١٠$^{-١٤}$ ٠,٩٩٩٩

(٢) ٠,٣١١٥ ، س̄ = ˘ ع^{٢} = ١,٥ ليتوفر شرط الإستقلال

(٣) ٠,٦٧٢

(٤) وفق توزيع ذى الحدين دليلاه ٣ ، ٠,٠٢ ثم قارن التكرارات المتوقعة بالتكرارات المشاهدة .

تمارين (٣ - ٢) :

(١) ٠,٤٨٦٨ ٠,٣٥٠٥ ٠,١٢٦٢ ٠,٠٣٦٥

(٢) ٠,٩٠٤٨ ٠,٠٠٤٧

(٣) ٠,٥٧٦٧

(٤) س̄ = ٠,٤٣٦٣ ، ع^{٢} = ٠,٩٧٠٩

٣٨٠,٧ ١٦٦,١ ٣٦,٢ ٥,٣٠ ٠,٦ ٠,١ ٠,٠ ٠,٠ ٠,٠

(٥) س̄ = ٠,٤٦٤٣ ، ع^{٢} = ٠,٢٦٩

التكررات المتوقعة ٧٠,٤ ٣٢,٧ ٧,٦ ١,٢ ٠,١

(٦) ٠,٣٦٧٩ ٠,٢٦٤٢

تمارين (٤) :

(١) - أ - ٠,٤٣٣٢ ٠,٠٤٩٥ ٠,٢٢٩٦ ٠,٨٧٤٢

- ب - ٠,١٥٥٤ ٠,٠٠٨٠

(٢) ,٩٨٢٠ ٠,٦٨٢٦

(٣) ٠,٩٩٩٩ ٠,٠٤٣٦

تمارين (٥) :

(١) - أ - ٢,١٧٩ ٢,٤٢٣ ٢,٩٢١ ١,٩٦٠

- ب - ٣١,٤١٠ ٥٧,٣٤ ٤٧,٢١٢ ٣,٨٤١

- جـ - ٣,٥٨ ٤,١٥ ٤,١٧ ٥,٧٥

(٢) – أ –   ٠,٠٥   ٠,٠٢٥
– ب –   ٠,٠١   ٠,٥
– جـ –   ٠,٠١   ٠,٠٢٥

تمارين (٦ ــ١) :

(١) ت × ٠,٠٠٥٥ نرفض الفرض الصفرى في الحالتين – (٠,٦٨٥٤ ، ٠,٩١٦٦)

(٢) ت = ٥,٦٩٣ نرفض الفرض الصفرى – (٨٢,٥٣ ، ٨٨,٠٧٢)

(٣) (٧,٩٩ ، ١٠,٨١)

(٤) ع = ٠,٦٩٢٨ خ . م = ٠,٤٣٧٩ ت = ــ ٢٧٤, الفرق ليس ذا دلالة

(٥) $ع^2$ = ١٧٢,٣٥٧خ . م = ٧,٠٥٨ ت = ــ ١,٠١٢ الفرق ليس ذا دلالة

(٦) نرفض الفرض الصفرى في الحالتين – (٤٢,٨٩ ، ٤٥,٧٨) – (٤٢,٤١ ، ٤٦,٢٥)

تمارين (٦ – ٢) :

(١) لايوجد دليل للشك في نظرية مندل χ^2 = ٠,١٣٧ ، χ^2 = ٠,٤٥٠

(٢) χ^2 = ١٠,٥٢٦ الفرق بين المصايد ذو دلالة عند المستوى ٠,٠٥

(٣) χ^2 = ١,٩٤ عدد المواليد غير ثابت خلال شهور السنة (لايعتمد الاختبار لأن هناك نمط) .

(٤) χ^2 = ٤,٣٢ نقبل الفرض الصفرى عند ٠,٠١ ونرفضه عند ٠,٠٥ المجتمع ربما يفضل النوع أ

(٥) χ^2 = ٥,٢٢٤ نقبل الفرض الصفرى أن الزهر غير متحيز عند المستوى ٠,٠٥

(٧) χ^2 = ٢٣,٦٨٣ حيوية الحبوب غير مستقلة عن المعالجة الحرارية .

(٨) χ^2 = ٢,٣٨ الدواء ليس له تأثير بناء على هذه التجربة .

(٩) (٤,٨٥ ، ١١١,٧٤)
(١٠) (٦٨٣,٩١ ، ١٧٠٠,٨٥)
(١١) $\chi^2_{ى}$ = ٩,٨٧

تمارين (٦ – ٣) :

(١) (٠,٢٩٤ ، ٠,٣٨٦)(٢) (٠,٠٨٧ ، ٠,٢٣٣)(٣) (٠,٣١٦ ، ٠,٧٧٠)
(٤) ص ى = – ٦,٢٨٤ نرفض فب (٥) ص ى = ٢,٩٤٨ نرفض ف

تمارين (٦ – ٤) :

(١) ن = ٦٨ (٢) (٣٠,٩٥ ، ٣٣,٠٥) ، ن = ٩٦, (٣) ن = ٥٤٥

تمارين (٦ – ٥) :

حدا المراقبة ٣،٧

تمارين (٧) :

(١) ا = ٤٨,٩٢ ، ا١ = ٤٢,٠٨ ، ه = ٠,٩٩٨
(٢) أولا : ا = ٢٥,٥٥٨ ، ه = ٠,٩٤
ثانيا : ا = ٢٥,٨٦٢ ، ا٢ = ٢٢,١٣٨ ، ه = ٠,٨٨ .
(٣) ه = ٠,١٣ تقريبا .

تمارين (٨ – ١)

	مصدر التباين	م ٢	د.ح.	تقدير التباين	ف
(١)	بين الأقسام	٦٦,١٧	٢	٣٣,٠٨٥	> ١
	داخل الأقسام	٥٠٨,٥٧	٩	٥٦,٥٣	
	الكلى	٥٧٤,٩٢	١١		
	$\mu_1 = \mu_2 = \mu_3$				
(٢)	بين الأقسام	١٨٠٧,٧٣	٣	٦٠٢,٥٧	٥,٢٦**

	داخل الأقسام	٣٧٧٨	٣٣	١١٤,٤٨	
	الكلى	٥٥٨٥,٧٣	٣٦		
(٣)	بين الأقسام	٠,٥٠٥	٢	٠,٢٥٢٥	١>
	داخل الأقسام	٤,٧٠٠	٩	٠,٥٢٢٢	
	الكلى	٥,٢٠٥	١١		

(٤) ف = ٥,٩٣ هناك دليل على وجود فروق بين المعالجات .

(٥) ف = ٣,٨٩٩ نرفض ف. ، تختلف أنواع الخرسانة فى متوسط امتصاصها للرطوبة .

(٦) (أولا) ف = ٢٣,٢٧٥نرفض القول بتساوى أطوال الدورات الثلاث .

(ثانيا) ف = ٣,٦٧٦ نقبل الفرض الصفرى عند ٠,٠٥ ونرفضه عند ٠,٠١ .

(٧) ع٢ = ٠,٠٣٩٣ (أ) هناك اختلاف جوهرى بين مجموعتى العلاج ومجموعة المراقبة .

(ب) ليس هناك خلاف جوهرى بين نوعى العلاج .

تمارين (٨ – ٢)

(١) أولا : ٣,٦٧ ، ٤,٢٢ كل من العاملين ذو دلالة عالية .

ثانيا : ٤,٢٦ نرفض عند ٠,٠٥ .

(٢) للغذاء ف = ٦,٦٢ متوسطات الكلوسترول ليست متساوية فى أنواع الغذاء .

للمعامل ف= ٤,٨٦ متوسطات المعامل ليست متساوية .

(٣) عامل الأيام ف = ٢,٣٠ ليست ذات دلالة .

عامل العمق ف = ٢٨٥١,١ ذودلالة عالية . درجة الحرارة تنخفض بزيادة العمق .

تمارين (٨ – ٣)

بين الأعمدة (السلالات)	٢,٧٢٧	١	٢,٧٢٧	< ١
بين الصفوف (الملوحة)	٩٨,٧٦١	٢	٤٩,٣٨١	٤,٦٩٩٩*
تفاعل	٢٤,٢١٥	٢	١٢,١٠٨	١,١٥٢٣
خطأ	١٨٩,١٢٤	١٨	١٠,٥٠٧	
كلى	٣١٤,٨٢٧	٢٣		

تمارين (٨ – ٤)

(١) بين القسمين (الأعمدة)	٥٨٩,٦٣	١	٥٨٩,٦٣	٨,٥٢٩
بين الأفراد (الصفوف)	٤٤٨,٨٧	١٤	٣٢,٠٦	,٤٦٤
الخطأ	٩٦٧,٨٧	١٤	٦٩,١٣	
الكلى	٢٠٠٦,٣٧	٢٩		

(٢) ميزان الطبيب يعطى قراءات أعلى – (٠,٠٤٢٨ ، ٢,٧٥٧٢) .

تمارين (٨ – ٥)

(١) بين الصفوف	١٧,٢٥	٣	٥,٧٥	٠,٩
بين الأعمدة	١١٤,٧٥	٣	٣٨,٢٥	٥,٩*
بين المعالجات	١٧٤,٧٥	٣	٥٨,٢٥	٩,٠*
الخطأ	٣٩,٠٠	٦		
الكلى	٣٤٥,٧٥	١٥		

تمارين (٨ – ٦)

(١) ثانيا :

مصدر التباين	م م	د ح	ط م	ف
بين الأقسام	٣٣١٢	٤	٨٢٨	٢٠,٧**
$\hat{\psi}_1$	٤٠٥	١	٤٠٥	١٠,١٢٥**
$\hat{\psi}_2$	٢٨٨٠	١	٢٨٨٠	٧٢**
$\hat{\psi}_3$	٢٢,٥	١	٢٢,٥	١>
$\hat{\psi}_4$	٤,٥	١	٤,٥	١>
الخطأ	١٨٠٠	٤٥	٤٠	
الكلى	٥١١٢	٤٩		

ثالثا : لأى مقارنة ع$^{2}_{\psi}$ = ٨ – القيمة الحرجة ١١,٠٧١

(٢) – أولا

مصدر التباين	م م	د خ	تقدير التباين	ف
بين الأعمدة	٣١١٤١,١٩	٣	١٠٣٨٠,٣٩٦	٣٣,٢٦٤
بين الصفوف	٤٦٠,٠٦	٢	٢٣٠,٠٣	١>
تفاعل	٣٤٩٩,٧٢	٦	٥٨٣,٢٨٧	١,٨٦٩
الخطأ	٧٤٨٩,٣٤	٢٤	٣١٢,٠٥٦	
الكلى	٤٢٥٩٠,٣١	٣٥		

ثانيا : متوسطات الأعمدة ٩٤,٣٣٣ ، ٩٧,٦٦٧ ، ١٤٥,٨٨٩ ، ١٦١,٥٥٦

ثالثا : للمقارنات البعدية للأعمدة ، القيمة الحرجة ٣١,٣٣٦

تمارين (٩ – ١)

(١) صَ = ١٢,٠٧ – ٧,٩٩ س ، عِ ص س = ٢,٣٥٨ ، ت = ٢,٧٨٥ ، (٢٢,١٢ ، ٣٤,٢٣)

(٢) صَ = ١,٩٣ س + ١٨٤,٦٦ ، ع ص س = ٥٩,٧٤١ ، ت = ٠,١٠٢ هناك علاقة خطية .

(٤) صَ = ٥,٢٨ + ٠,٠٠١٩ س ، ع ص س = ٠,٩٩٥ ، ت = ٠,٩٤٥ لا توجد علاقة خطية .

(٦) ت = ٠,١٤٧ لا توجد علاقة خطية .

تمارين (٩ – ٢)

(١) صَ = ٢٧,٨٦ – ٠,٥٩٩٩٦ س ، ف ى = ٢٥٧,٤٤ الانحراف عن الخطية ذو دلالة .

ف ى = ٩٥,٨٣ نرفض β = ٠ . هناك علاقة خطية ولكنها ليست أحسن العلاقات .

(٢) العلاقة الخطية تعبر تعبيرا جيدا عن العلاقة الحقيقية بين المتغيرين .

(٤) ف = ٨,٧٥ ص ليست مستقلة عن س .

ف = ٩,١٥ يوجد انحراف عن الخطية .

ف = ١٠,٠٠ هناك انحدار خطى ولكنه ليس أفضل العلاقات .

تمارين (١٠)

(١) ر = ٠,٤١٨ ، ت = ١,٤٥٥ لا توجد علاقة خطية .

(٢) ر = ٠,٨٩٨٣ ، ت = ١١,١١٧٧ توجد علاقة خطية .
(٤) ر = ٠,١٢٨ ، ت = ٣٦٥, لا توجد علاقة خطية .
(٥) ر = −٠,٩٧٨٦ ، ت = ١٢,٦ توجد علاقة خطية سالبة .
(٦) ر_ر = ٠,٧٩٤ ، القيمة الحرجة عند ن = ١٠ والمستوى ٠,٠١ هى ٠,٧٤٥ . هناك ارتباط موجب
(٧) ر_ر = ٠,٩٨ ، هناك ارتباط موجب .
(٨) معاملا الارتباط متساويان فى المجتمع .

تمارين (١١ − ١)

ف_ى = ٣٦,٠٣ يوجد انحدار خطى ، ف = ٢,١٤ المتوسطات متساوية فى المجتمع .

المصدر	ب	حـ²/ا	ب − حـ²/ا	د ح	ط م	ف
بين الأقسام	—	—	٦٨,٦٩٧	٢	٣٤,٣٥	٢,١٤
داخل الأقسام	٩٩٥,١	٥٧٨	٤١٧,١٠٠	٢٦	١٦,٠٤	
الكلى	١٢٨٨,٧٠٠	٨٠٢,٩٠٣	٤٨٥,٧٩٧	٢٨		

تمارين (١١ − ٢)

ف_ص = ٢٤,٤ يوجد تأثير خطى ، ف = ٤,٧٩ المتوسطات ليست متساوية .
ب = − ٠,٣٤ المتوسطات المعدلة ٢٣,٧٧ ، ٢٧,٥٢ ، ٢٥,١٩ ، ٢٤,٨٧ .

تمارين (١٢ - ١)

$\hat{ص}$ = ٩,٧٨٥٦ + ٠,٠٠١٧ $س_{١}$ + ٠,٢١٧٤ $س_{٢}$

ف = ٥,٨٣١٩ نرفض $\beta_{١} = \beta_{٢} = ٠$

$ت_{١}$ = ٢,١٠ نقبل $\beta_{١} = ٠$ ، $ت_{٢}$ = ٢,٨٧ نرفض $\beta_{٢} = ٠$

تمارين (١٢ - ٢)

(١) ف = ٤,٤٩٩٦ نرفض $\rho = ٠$

(٢) $r_{١٣-٢}$ = ٠,٤٤٥ ، ت = ٢,٩٨ نرفض $\rho_{١٣-٢} = ٠$

تمارين (١٤-١)

(١) $\bar{س}$ = ١٠ ، $\mu_{س}$ = ١٦ ، $\sigma_{س}$ = ٢,٦٠٣ ، ص = ٢,١١٣ .

هناك نمط دورى

(٢) الوسيط = ٣٠,٥ ، $\bar{س}$ = ١٣ ، $\mu_{س}$ = ١٣ ، $\sigma_{س}$ = ٢,٣٩٦

، ص = ± ٠,٢٠٩ . لا يوجد دليل ضد الفرض أن العينة عشوائية .

(٣) الوسيط = ٣٥ ، $\bar{س}$ = ١٠ ، $\mu_{س}$ = ١١ ، $\sigma_{س}$ = ٢,١٧٦ ،

ص = — ٠,٢٣ . لا يوجد دليل ضد الفرض أن العينة عشوائية .

تمارين (١٤ - ٢)

(١) ن = ١٥ ، س = ٤ ، ل (س ≤ ٤) = ٠,٠٥٦ .

نقبل أن المتوسط يساوى ٠,٥٥ .

(٢) ن = ٢٦ ، س = ٦ ، ص = -٢٢,٥٥ .

نرفض الفرض الصفرى . العالم الأول أفضل .

تمارين (١٤ - ٣)

ن = ٥ ، س = ١ الامتناع عن التدخين يزيد الوزن .

تمارين (١٤ - ٤)

(١) ن = ٧ ، ع = ٣ الأوكسجين ينقص مع العمق

(٢) ن = ١٠ ، ع = ٧ وزن القلب يزداد بازدياد ضغط الدم .

تمارين (١٤ - ٥)

(١) $ص_ى$ = ٥,٥٨ نرفض ف، عند α = ٠,١٠ ويبدو أن مستوى التلوث أكبر في النهر الثاني .

(٢) $ص_ى$ = ٩,٢٠ نرفض ف، عند α = ٠,١٠ اللون له تأثير .

ملحق (٢)
جداول احصائية

الجدول (١)
٢٥٠٠ من الأرقام العشوائية

	1	2	3	4	5	6	7	8	9	10	
1	48461	14952	72619	73689	52059	37086	60050	86192	67049	64739	1
2	76534	38149	49692	31366	52093	15422	20498	33901	10319	43397	2
3	70437	25861	38504	14752	23757	59660	67844	78815	23758	86814	3
4	59584	03370	42806	11393	71722	93804	09095	07856	55589	46020	4
5	04285	58554	16085	51555	27501	73883	33427	33343	45507	50063	5
6	77340	10412	69189	85171	29082	44785	83638	02583	96483	76553	6
7	59183	62687	91778	80354	23512	97219	65921	02035	59847	91403	7
8	91800	04281	39979	03927	82564	28777	59049	97532	54540	79472	8
9	12066	24817	81099	48940	69554	55925	48379	12866	51232	21580	9
10	69907	91751	53512	23748	65906	91385	84983	27915	48491	91068	10
11	80467	04873	54053	25955	48518	13815	37707	68687	15570	08890	11
12	78057	67835	28302	45048	56761	97725	58438	91528	24645	18544	12
13	05648	39387	78191	88415	60269	94880	58812	42931	71898	61534	13
14	22304	39246	01350	99451	61862	78688	30339	60222	74052	25740	14
15	61346	50269	67005	40442	33100	16742	61640	21046	31909	72641	15
16	66793	37696	27965	30459	91011	51426	31006	77468	61029	57108	16
17	86411	48809	36698	42453	83061	43769	39948	87031	30767	13953	17
18	62098	12825	81744	28882	27369	88183	65846	92545	09065	22655	18
19	68775	06261	54265	16203	23340	84750	16317	88686	86842	00879	19
20	52679	19595	13687	74872	89181	01939	18447	10787	76246	80072	20
21	84096	87152	20719	25215	04349	54434	72344	93008	83282	31670	21
22	63964	55937	21417	49944	38356	98404	14850	17994	17161	98981	22
23	31191	75131	72386	11689	95727	05414	88727	45583	22568	77700	23
24	30545	68523	29850	67833	05622	89975	79042	27142	99257	32349	24
25	52573	91001	52315	26430	54175	30122	31796	98842	37600	26025	25
26	16586	81842	01076	99414	31574	94719	34656	80018	86988	79234	26
27	81841	88481	61191	25013	30272	23388	22463	65774	10029	58376	27
28	43563	66829	72838	08074	57080	15446	11034	98143	74989	26885	28
29	19945	84193	57581	77252	85604	45412	43556	27518	90572	00563	29
30	79374	23796	16919	99691	80276	32818	62953	78831	54395	30705	30
31	48503	26615	43980	09810	38289	66679	73799	48418	12647	40044	31
32	32049	65541	37937	41105	70106	89706	40829	40789	59547	00783	32
33	18547	71562	95493	34112	76895	46766	96395	31718	48302	45893	33
34	03180	96742	61486	43305	34183	99605	67803	13491	09249	29557	34
35	94822	24738	67749	83748	59799	25210	31093	62925	72061	69991	35
36	34330	60599	85828	19152	68499	27977	35611	96240	62747	89529	36
37	43770	81537	59527	95674	76692	86420	69930	10020	72881	12532	37
38	56908	77192	50623	41215	14311	42834	80651	93750	59957	31211	38
39	32787	07189	80539	75927	75475	73965	11796	72140	48944	74156	39
40	52441	78392	11733	57703	29133	71164	55355	31006	25526	55790	40
41	22377	54723	18227	28449	04570	18882	00023	67101	06895	08915	41
42	18376	73460	88841	39602	34049	20589	05701	08249	74213	25220	42
43	53201	28610	87957	21497	64729	64983	71551	99016	87903	63875	43
44	34919	78901	59710	27396	07593	05665	11964	44134	00273	76358	44
45	33617	92159	21971	16901	57383	34262	41744	60891	57624	06962	45
46	70010	40964	98780	72418	52571	18415	64362	90636	38034	04909	46
47	19282	68447	35665	31530	59832	49181	21914	65742	89815	39231	47
48	91429	73328	13266	54898	68795	40948	80808	63887	89939	47938	48
49	97637	78393	33021	05867	86520	45363	43066	00988	64040	09803	49
50	95150	07625	05255	83254	93943	52325	93230	62668	79529	65964	50

الجدول (٢)

معاملات ذى الحدين : ن قر

k	$\binom{1}{k}$	$\binom{2}{k}$	$\binom{3}{k}$	$\binom{4}{k}$	$\binom{5}{k}$	$\binom{6}{k}$	$\binom{7}{k}$	$\binom{8}{k}$	$\binom{9}{k}$	$\binom{10}{k}$	$\binom{11}{k}$	$\binom{12}{k}$	$\binom{13}{k}$
0	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	2	3	4	5	6	7	8	9	10	11	12	13
2		1	3	6	10	15	21	28	36	45	55	66	78
3			1	4	10	20	35	56	84	120	165	220	286
4				1	5	15	35	70	126	210	330	495	715
5					1	6	21	56	126	252	462	792	1287
6						1	7	28	84	210	462	924	1716
7							1	8	36	120	330	792	1716
8								1	9	45	165	495	1287
9									1	10	55	220	715
10										1	11	66	286
11											1	12	78
12												1	13
13													1

k	$\binom{14}{k}$	$\binom{15}{k}$	$\binom{16}{k}$	$\binom{17}{k}$	$\binom{18}{k}$	$\binom{19}{k}$	$\binom{20}{k}$
0	1	1	1	1	1	1	1
1	14	15	16	17	18	19	20
2	91	105	120	136	153	171	190
3	364	455	560	680	816	969	1140
4	1001	1365	1820	2380	3060	3876	4845
5	2002	3003	4368	6188	8568	11628	15504
6	3003	5005	8008	12376	18564	27132	38760
7	3432	6435	11440	19448	31824	50388	77520
8	3003	6435	12870	24310	43758	75582	125970
9	2002	5005	11440	24310	48620	92378	167960
10	1001	3003	8008	19448	43758	92378	184756
11	364	1365	4368	12376	31824	75582	167960
12	91	455	1820	6188	18564	50388	125970
13	14	105	560	2380	8568	27132	77520
14	1	15	120	680	3060	11628	38760
15		1	16	136	816	3876	15504
16			1	17	153	969	4845
17				1	18	171	1140
18					1	19	190
19						1	20
20							1

تابع الجدول (٣)

الاحتمالات في توزيع ذي الحدين : د (س)

n	x	p = 0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
9	0	0.630	0.387	0.134	0.040	0.010	0.002					
	1	0.299	0.387	0.302	0.156	0.060	0.018	0.004				
	2	0.063	0.172	0.302	0.267	0.161	0.070	0.021	0.004			
	3	0.008	0.045	0.176	0.267	0.251	0.164	0.074	0.021	0.003		
	4	0.001	0.007	0.066	0.172	0.251	0.246	0.167	0.074	0.017	0.001	
	5		0.001	0.017	0.074	0.167	0.246	0.251	0.172	0.066	0.007	0.001
	6			0.003	0.021	0.074	0.164	0.251	0.267	0.176	0.045	0.008
	7				0.004	0.021	0.070	0.161	0.267	0.302	0.172	0.063
	8					0.004	0.018	0.060	0.156	0.302	0.387	0.299
	9						0.002	0.010	0.040	0.134	0.387	0.630
10	0	0.599	0.349	0.107	0.028	0.006	0.001					
	1	0.315	0.387	0.268	0.121	0.040	0.010	0.002				
	2	0.075	0.194	0.302	0.233	0.121	0.044	0.011	0.001			
	3	0.010	0.057	0.201	0.267	0.215	0.117	0.042	0.009	0.001		
	4	0.001	0.011	0.088	0.200	0.251	0.205	0.111	0.037	0.006		
	5		0.001	0.026	0.103	0.201	0.246	0.201	0.103	0.026	0.001	
	6			0.006	0.037	0.111	0.205	0.251	0.200	0.088	0.011	0.001
	7			0.001	0.009	0.042	0.117	0.215	0.267	0.201	0.057	0.010
	8				0.001	0.011	0.044	0.121	0.233	0.302	0.194	0.075
	9					0.002	0.010	0.040	0.121	0.268	0.387	0.315
	10						0.001	0.006	0.028	0.107	0.349	0.599
11	0	0.569	0.314	0.086	0.020	0.004						
	1	0.329	0.384	0.236	0.093	0.027	0.005	0.001				
	2	0.087	0.213	0.295	0.200	0.089	0.027	0.005	0.001			
	3	0.014	0.071	0.221	0.257	0.177	0.081	0.023	0.004			
	4	0.001	0.016	0.111	0.220	0.236	0.161	0.070	0.017	0.002		
	5		0.002	0.039	0.132	0.221	0.226	0.147	0.057	0.010		
	6			0.010	0.057	0.147	0.226	0.221	0.132	0.039	[illegible]	
	7			0.002	0.017	0.070	0.161	0.236	0.220	0.111	0.016	0.001
	8				0.004	0.023	0.081	0.177	0.257	0.221	0.071	0.014
	9				0.001	0.005	0.027	0.089	0.200	0.295	0.213	0.087
	10					0.001	0.005	0.027	0.093	0.236	0.384	0.329
	11							0.004	0.020	0.086	0.314	0.569
12	0	0.540	0.282	0.069	0.014	0.002						
	1	0.341	0.377	0.206	0.071	0.017	0.003					
	2	0.099	0.230	0.283	0.168	0.064	0.016	0.002				
	3	0.017	0.085	0.236	0.240	0.142	0.054	0.012	0.001			
	4	0.002	0.021	0.133	0.231	0.213	0.121	0.042	0.008	0.001		
	5		0.[illegible]	0.053	0.158	0.227	0.193	0.101	0.029	0.003		
	6			0.016	0.079	0.177	0.226	0.177	0.079	0.016		
	7			0.003	0.029	0.101	0.193	0.227	0.158	0.053	0.004	
	8			0.001	0.008	0.042	0.121	0.213	0.231	0.133	0.021	0.002
	9				0.001	0.012	0.054	0.142	0.240	0.236	0.085	0.017
	10					0.002	0.016	0.064	0.168	0.283	0.230	0.099
	11						0.003	0.017	0.071	0.206	0.377	0.341
	12							0.002	0.014	0.069	0.282	0.540

الجدول (٣)

الاحتمالات في توزيع ذى الحدين : د (س)

n	x	0.05	0.1	0.2	[illegible]	0.4	0.5	0.6	0.7	0.8	0.9	0.95
2	0	0.902	0.810	0.640	0.490	0.360	0.250	0.160	0.090	0.040	0.010	0.002
	1	0.095	0.180	0.320	0.420	0.480	0.500	0.480	0.420	0.320	0.180	0.095
	2	0.002	0.010	0.040	0.090	0.160	0.250	0.360	0.490	0.640	0.810	0.902
3	0	0.857	0.729	0.512	0.343	0.216	0.125	0.064	0.027	0.008	0.001	
	1	0.135	0.243	0.384	0.441	0.432	0.375	0.288	0.189	0.096	0.027	0.007
	2	0.007	0.027	0.096	0.189	0.288	0.375	0.432	0.441	0.384	0.243	0.135
	3		0.001	0.008	0.027	0.064	0.125	0.216	0.343	0.512	0.729	0.857
4	0	0.815	0.656	0.410	0.240	0.130	0.062	0.026	0.008	0.002		
	1	0.171	0.292	0.410	[illegible]	0.346	0.250	0.154	0.076	0.026	0.004	
	2	0.014	0.049	0.154	0.265	0.346	0.375	0.346	0.265	0.154	0.049	0.014
	3		0.004	0.026	0.076	0.154	0.250	0.346	0.412	0.410	0.292	0.171
	4			0.002	0.008	0.026	0.062	0.130	0.240	0.410	0.656	0.815
5	0	0.774	0.590	0.328	0.168	0.078	0.031	0.010	0.002			
	1	0.204	0.328	0.410	0.360	0.259	0.156	0.077	0.028	0.006		
	2	0.021	0.073	0.205	0.309	0.346	0.312	0.230	0.132	0.051	0.008	0.001
	3	0.001	0.008	0.051	0.132	0.230	0.312	0.346	0.309	0.205	0.073	0.021
	4			0.006	0.028	0.077	0.156	0.259	0.360	0.410	0.328	0.204
	5				0.002	0.010	0.031	0.078	0.168	0.328	0.590	0.774
6	0	0.735	0.531	0.262	0.118	0.047	0.016	0.004	0.001			
	1	0.232	0.354	0.393	0.303	0.187	0.094	0.037	0.010	0.002		
	2	[illegible]	0.098	0.246	0.324	0.311	0.234	0.138	0.060	0.015	[illegible]	
	3	0.002	0.015	0.082	0.185	0.276	0.312	0.276	0.185	0.082	0.015	0.002
	4		0.001	0.015	0.060	0.138	0.234	0.311	0.324	0.246	0.098	0.031
	5			0.002	0.010	0.037	0.094	0.187	0.303	0.393	0.354	0.232
	6				0.001	0.004	0.016	0.047	0.118	0.262	0.531	0.735
7	0	0.698	0.478	0.210	0.082	0.028	0.008	0.002				
	1	0.257	0.372	0.367	0.247	0.131	0.055	0.017	0.004			
	2	0.041	0.124	0.275	0.318	0.261	0.164	0.077	0.025	0.004		
	3	0.004	0.023	0.115	0.227	0.290	0.273	0.194	0.097	0.029	0.003	
	4		0.003	[illegible]	0.097	0.194	0.273	0.290	0.227	0.115	0.023	0.004
	5			0.004	0.025	0.077	0.164	0.261	0.318	0.275	0.124	0.041
	6				0.004	0.017	0.055	0.131	0.247	0.367	0.372	0.257
	7					0.002	0.008	0.028	0.082	0.210	0.478	0.698
8	0	0.663	0.430	0.168	0.058	0.017	0.004	0.001				
	1	0.279	0.383	0.336	0.198	0.090	0.031	0.008	0.001			
	2	0.051	0.149	0.294	0.296	0.209	0.109	0.041	0.010	0.001		
	3	0.005	0.033	0.147	0.254	0.279	0.219	0.124	0.047	0.009		
	4		0.005	0.046	0.136	0.232	0.273	0.232	0.136	0.046	0.005	
	5			0.009	0.047	0.124	0.219	0.279	0.254	0.147	0.033	0.005
	6			0.001	0.010	0.041	0.109	0.209	0.296	0.294	0.149	0.051
	7				0.001	0.008	0.031	0.090	0.198	0.336	0.383	0.279
	8					0.001	0.004	0.017	0.058	0.168	0.430	0.663

All values omitted in this table are 0.0005 or less

تابع الجدول (٣)

الاحتمالات في توزيع ذي الحدين : د (س)

n	x	p 0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
13	0	0.513	0.254	0.055	0.010	0.001						
	1	0.351	0.367	0.179	0.054	0.011	0.002					
	2	0.111	0.245	0.268	0.139	0.045	0.010	0.001				
	3	0.021	0.100	0.246	0.218	0.111	0.035	0.006	0.001			
	4	0.003	0.028	0.154	0.234	0.184	0.087	0.024	0.003			
	5		0.006	0.069	0.180	0.221	0.157	0.066	0.014	0.001		
	6		0.001	0.023	0.103	0.197	0.209	0.131	0.044	0.006		
	7			0.006	0.044	0.131	0.209	0.197	0.103	0.023	0.001	
	8			0.001	0.014	0.066	0.157	0.221	0.180	0.069	0.006	
	9				0.003	0.024	0.087	0.184	0.234	0.154	0.028	0.003
	10				0.001	0.006	0.035	0.111	0.218	0.246	0.100	0.021
	11					0.001	0.010	0.045	0.139	0.268	0.245	0.111
	12						0.002	0.011	0.054	0.179	0.367	0.351
	13							0.001	0.010	0.055	0.254	0.513
14	0	0.488	0.229	0.044	0.007	0.001						
	1	0.359	0.356	0.154	0.041	0.007	0.001					
	2	0.123	0.257	0.250	0.113	0.032	0.006	0.001				
	3	0.026	0.114	0.250	0.194	0.085	0.022	0.003				
	4	0.004	0.035	0.172	0.229	0.155	0.061	0.014	0.001			
	5		0.008	0.086	0.196	0.207	0.122	0.041	0.007			
	6		0.001	0.032	0.126	0.207	0.183	0.092	0.023	0.002		
	7			0.009	0.062	0.157	0.209	0.157	0.062	0.009		
	8			0.002	0.023	0.092	0.183	0.207	0.126	0.032	0.001	
	9				0.007	0.041	0.122	0.207	0.196	0.086	0.008	
	10				0.001	0.014	0.061	0.155	0.229	0.172	0.035	0.004
	11					0.003	0.022	0.085	0.194	0.250	0.114	0.026
	12					0.001	0.006	0.032	0.113	0.250	0.257	0.123
	13						0.001	0.007	0.041	0.154	0.356	0.359
	14							0.001	0.007	0.044	0.229	0.488
15	0	0.463	0.206	0.035	0.005							
	1	0.366	0.343	0.132	0.031	0.005						
	2	0.135	0.267	0.231	0.092	0.022	0.003					
	3	0.031	0.129	0.250	0.170	0.063	0.014	0.002				
	4	0.005	0.043	0.188	0.219	0.127	0.042	0.007	0.001			
	5	0.001	0.010	0.103	0.206	0.186	0.092	0.024	0.003			
	6		0.002	0.043	0.147	0.207	0.153	0.061	0.012	0.001		
	7			0.014	0.081	0.177	0.196	0.118	0.035	0.003		
	8			0.003	0.035	0.118	0.196	0.177	0.081	0.014		
	9			0.001	0.012	0.061	0.153	0.207	0.147	0.043	0.002	
	10				0.003	0.024	0.092	0.186	0.206	0.103	0.010	0.001
	11				0.001	0.007	0.042	0.127	0.219	0.188	0.043	0.005
	12					0.002	0.014	0.063	0.170	0.250	0.129	0.031
	13						0.003	0.022	0.092	0.231	0.267	0.135
	14							0.005	0.031	0.132	0.343	0.366
	15								0.005	0.035	0.206	0.463

الجدول (٤)

قيم هـ-م

λ	1	2	3	4	5	6	7	8	9	10
$e^{-\lambda}$	0.36788	0.13534	0.04979	0.01832	0.006738	0.002479	0.000912	0.000335	0.000123	0.000045

λ	0	1	2	3	4	5	6	7	8	9
0.0	1.0000	0.9900	0.9802	0.9704	0.9608	0.9512	0.9418	0.9324	0.9231	0.9139
0.1	0.9048	0.8958	0.8869	0.8781	0.8694	0.8607	0.8521	0.8437	0.8353	0.8270
0.2	0.8187	0.8106	0.8025	0.7945	0.7866	0.7788	0.7711	0.7634	0.7558	0.7483
0.3	0.7408	0.7334.	0.7261	0.7189	0.7118	0.7047	0.6977	0.6907	0.6839	0.6771
0.4	0.6703	0.6636	0.6570	0.6505	0.6440	0.6376	0.6313	0.6250	0.6188	0.6126
0.5	0.6065	0.6005	0.5945	0.5886	0.5827	0.5770	0.5712	0.5655	0.5599	0.5543
0.6	0.5488	0.5434	0.5379	0.5326	0.5273	0.5220	0.5169	0.5117	0.5066	0.5016
0.7	0.4966	0.4916	0.4868	0.4819	0.4771	0.4724	0.4677	0.4630	0.4584	0.4538
0.8	0.4493	0.4449	0.4404	0.4360	0.4317	0.4274	0.4232	0.4190	0.4148	0.4107
0.9	0.4066	0.4025	0.3985	0.3946	0.3906	0.3867	0.3829	0.3791	0.3753	0.3716

مثال : هـ-٢٫٣٤ = هـ-٢ . هـ-٠٫٣٤ = ٠٫١٣٥٣٤ × ٠٫٧١١٨

= ٠٫٠٩٦٣

الجدول (٥)

الاحتمالات والاحتمالات المتجمعة في توزيع بواسون

x	$\mu = 0.1$		$\mu = 0.2$		$\mu = 0.3$		$\mu = 0.4$		$\mu = 0.5$	
	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$
0	0.9048	0.9048	0.8187	0.8187	0.7408	0.7408	0.6703	0.6703	0.6065	0.6065
1	0905	0.9953	1637	0.9825	2222	0.9631	2681	0.9384	3033	0.9098
2	0045	0.9998	0164	0.9989	0333	0.9964	0536	0.9921	0758	0.9856
3	0002	1.0000	0011	0.9999	0033	0.9997	0072	0.9992	0126	0.9982
4	0000	1.0000	0001	1.0000	0003	1.0000	0007	0.9999	0016	0.9998
5							0001	1.0000	0002	1.0000

x	$\mu = 0.6$		$\mu = 0.7$		$\mu = 0.8$		$\mu = 0.9$		$\mu = 1$	
	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$
0	0.5488	0.5488	0.4966	0.4966	0.4493	0.4493	0.4066	0.4066	0.3679	0.3679
1	3293	0.8781	3476	0.8442	3595	0.8088	3659	0.7725	3679	0.7358
2	0988	0.9769	1217	0.9659	1438	0.9526	1647	0.9371	1839	0.9197
3	0198	0.9966	0284	0.9942	0383	0.9909	0494	0.9865	0613	0.9810
4	0030	0.9996	0050	0.9992	0077	0.9986	0111	0.9977	0153	0.9963
5	0004	1.0000	0007	0.9999	0012	0.9998	0020	0.9997	0031	0.9994
6			0001	1.0000	0002	1.0000	0003	1.0000	0005	0.9999
7									0001	1.0000

x	$\mu = 1.5$		$\mu = 2$		$\mu = 3$		$\mu = 4$		$\mu = 5$	
	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$
0	0.2231	0.2231	0.1353	0.1353	0.0498	0.0498	0.0183	0.0183	0.0067	0.0067
1	3347	0.5578	2707	0.4060	1494	0.1991	0733	0.0916	0337	0.0404
2	2510	0.8088	2707	0.6767	2240	0.4232	1465	0.2381	0842	0.1247
3	1255	0.9344	1804	0.8571	2240	0.6472	1954	0.4335	1404	0.2650
4	0471	0.9814	0902	0.9473	1680	0.8153	1954	0.6288	1755	0.4405
5	0141	0.9955	0361	0.9834	1008	0.9161	1563	0.7851	1755	0.6160
6	0035	0.9991	0120	0.9955	0504	0.9665	1042	0.8893	1462	0.7622
7	0008	0.9998	0034	0.9989	0216	0.9881	0595	0.9489	1044	0.8666
8	0001	1.0000	0009	0.9998	0081	0.9962	0298	0.9786	0653	0.9319
9			0002	1.0000	0027	0.9989	0132	0.9919	0363	0.9682
10					0008	0.9997	0053	0.9972	0181	0.9863
11					0002	0.9999	0019	0.9991	0082	0.9945
12					0001	1.0000	0006	0.9997	0034	0.9980
13							0002	0.9999	0013	0.9993
14							0001	1.0000	0005	0.9998
15									0002	0.9999
16									0000	1.0000

الجدول (٦)

المساحات أسفل المنحنى المعتدل المعيارى

Standard deviation units	0	1	2	3	4	5	6	7	8	9	Standard deviation units
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359	0.0
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753	0.1
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141	0.2
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517	0.3
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879	0.4
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224	0.5
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549	0.6
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852	0.7
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133	0.8
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389	0.9
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621	1.0
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830	1.1
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015	1.2
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177	1.3
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319	1.4
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441	1.5
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545	1.6
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633	1.7
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706	1.8
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767	1.9
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817	2.0
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857	2.1
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890	2.2
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916	2.3
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936	2.4
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952	2.5
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964	2.6
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974	2.7
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981	2.8
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986	2.9
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990	3.0
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993	3.1
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995	3.2
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997	3.3
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998	3.4
3.5	.499767										
3.6	.499841										
3.7	.499892										
3.8	.499928										
3.9	.499952										
4.0	.499968										
4.1	.499979										
4.2	.499987										
4.3	.499991										
4.4	.499995										
4.5	.499997										
4.6	.499998										
4.7	.499999										
4.8	.499999										
4.9	.500000										

Tabled area

α

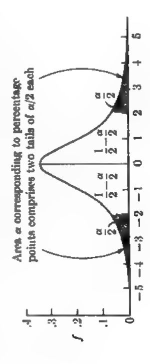
Argument $= \frac{Y - \mu_Y}{\sigma}$

Note: The quantity given is the area under the standard normal density function between the mean and the critical point. The area is generally labeled $\frac{1}{2} - \alpha$ (as shown in the figure). By inverse interpolation one can find the number of standard deviations corresponding to a given area.

الجدول (٧)

القيم الحرجة لتوزيع ت

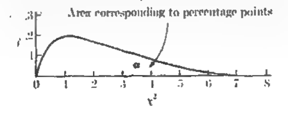
ν \ α	0.9	0.5	0.4	0.2	0.1	0.05	0.02	0.01	0.001	α / ν
1	.158	1.000	1.376	3.078	6.314	12.706	31.821	63.657	636.619	1
2	.142	.816	1.061	1.886	2.920	4.303	6.965	9.925	31.598	2
3	.137	.765	.978	1.638	2.353	3.182	4.541	5.841	12.924	3
4	.134	.741	.941	1.533	2.132	2.776	3.747	4.604	8.610	4
5	.132	.727	.920	1.476	2.015	2.571	3.365	4.032	6.869	5
6	.131	.718	.906	1.440	1.943	2.447	3.143	3.707	5.959	6
7	.130	.711	.896	1.415	1.895	2.365	2.998	3.499	5.408	7
8	.130	.706	.889	1.397	1.860	2.306	2.896	3.355	5.041	8
9	.129	.703	.883	1.383	1.833	2.262	2.821	3.250	4.781	9
10	.129	.700	.879	1.372	1.812	2.228	2.764	3.169	4.587	10
11	.129	.697	.876	1.363	1.796	2.201	2.718	3.106	4.437	11
12	.128	.695	.873	1.356	1.782	2.179	2.681	3.055	4.318	12
13	.128	.694	.870	1.350	1.771	2.160	2.650	3.012	4.221	13
14	.128	.692	.868	1.345	1.761	2.145	2.624	2.977	4.140	14
15	.128	.691	.866	1.341	1.753	2.131	2.602	2.947	4.073	15
16	.128	.690	.865	1.337	1.746	2.120	2.583	2.921	4.015	16
17	.128	.689	.863	1.333	1.740	2.110	2.567	2.898	3.965	17
18	.127	.688	.862	1.330	1.734	2.101	2.552	2.878	3.922	18
19	.127	.688	.861	1.328	1.729	2.093	2.539	2.861	3.883	19
20	.127	.687	.860	1.325	1.725	2.086	2.528	2.845	3.850	20
21	.127	.686	.859	1.323	1.721	2.080	2.518	2.831	3.819	21
22	.127	.686	.858	1.321	1.717	2.074	2.508	2.819	3.792	22
23	.127	.685	.858	1.319	1.714	2.069	2.500	2.807	3.767	23
24	.127	.685	.857	1.318	1.711	2.064	2.492	2.797	3.745	24
25	.127	.684	.856	1.316	1.708	2.060	2.485	2.787	3.725	25
26	.127	.684	.856	1.315	1.706	2.056	2.479	2.779	3.707	26
27	.127	.684	.855	1.314	1.703	2.052	2.473	2.771	3.690	27
28	.127	.683	.855	1.313	1.701	2.048	2.467	2.763	3.674	28
29	.127	.683	.854	1.311	1.699	2.045	2.462	2.756	3.659	29
30	.127	.683	.854	1.310	1.697	2.042	2.457	2.750	3.646	30
40	.126	.681	.851	1.303	1.684	2.021	2.423	2.704	3.551	40
60	.126	.679	.848	1.296	1.671	2.000	2.390	2.660	3.460	60
120	.126	.677	.845	1.289	1.658	1.980	2.358	2.617	3.373	120
∞	.126	.674	.842	1.282	1.645	1.960	2.326	2.576	3.291	∞



الجدول (٨)

القيم الحرجة لتوزيع χ^2

$\nu \backslash \alpha$	0.995	0.975	0.9	0.5	0.1	0.05	0.025	0.01	0.005	ν
1	.000	.000	0.016	0.455	2.706	3.841	5.024	6.635	7.879	1
2	0.010	0.051	0.211	1.386	4.605	5.991	7.378	9.210	10.597	2
3	0.072	0.216	0.584	2.366	6.251	7.815	9.348	11.345	12.838	3
4	0.207	0.484	1.064	3.357	7.779	9.488	11.143	13.277	14.860	4
5	0.412	0.831	1.610	4.351	9.236	11.070	12.832	15.086	16.750	5
6	0.676	1.237	2.204	5.348	10.645	12.592	14.449	16.812	18.548	6
7	0.989	1.690	2.833	6.346	12.017	14.067	16.013	18.475	20.278	7
8	1.344	2.180	3.490	7.344	13.362	15.507	17.535	20.090	21.955	8
9	1.735	2.700	4.168	8.343	14.684	16.919	19.023	21.666	23.589	9
10	2.156	3.247	4.865	9.342	15.987	18.307	20.483	23.209	25.188	10
11	2.603	3.816	5.578	10.341	17.275	19.675	21.920	24.725	26.757	11
12	3.074	4.404	6.304	11.340	18.549	21.026	23.337	26.217	28.300	12
13	3.565	5.009	7.042	12.340	19.812	22.362	24.736	27.688	29.819	13
14	4.075	5.629	7.790	13.339	21.064	23.685	26.119	29.141	31.319	14
15	4.601	6.262	8.547	14.339	22.307	24.996	27.488	30.578	32.801	15
16	5.142	6.908	9.312	15.338	23.542	26.296	28.845	32.000	34.267	16
17	5.697	7.564	10.085	16.338	24.769	27.587	30.191	33.409	35.718	17
18	6.265	8.231	10.865	17.338	25.989	28.869	31.526	34.805	37.156	18
19	6.844	8.907	11.651	18.338	27.204	30.144	32.852	36.191	38.582	19
20	7.434	9.591	12.443	19.337	28.412	31.410	34.170	37.566	39.997	20
21	8.034	10.283	13.240	20.337	29.615	32.670	35.479	38.932	41.401	21
22	8.643	10.982	14.042	21.337	30.813	33.924	36.781	40.289	42.796	22
23	9.260	11.688	14.848	22.337	32.007	35.172	38.076	41.638	44.181	23
24	9.886	12.401	15.659	23.337	33.196	36.415	39.364	42.980	45.558	24
25	10.520	13.120	16.473	24.337	34.382	37.652	40.646	44.314	46.928	25
26	11.160	13.844	17.292	25.336	35.563	38.885	41.923	45.642	48.290	26
27	11.808	14.573	18.114	26.336	36.741	40.113	43.194	46.963	49.645	27
28	12.461	15.308	18.939	27.336	37.916	41.337	44.461	48.278	50.993	28
29	13.121	16.047	19.768	28.336	39.088	42.557	45.722	49.588	52.336	29
30	13.787	16.791	20.599	29.336	40.256	43.773	46.979	50.892	53.672	30
31	14.458	17.539	21.434	30.336	41.422	44.985	48.232	52.191	55.003	31
32	15.134	18.291	22.271	31.336	42.585	46.194	49.480	53.486	56.329	32
33	15.815	19.047	23.110	32.336	43.745	47.400	50.725	54.776	57.649	33
34	16.501	19.806	23.952	33.336	44.903	48.602	51.966	56.061	58.964	34
35	17.192	20.569	24.797	34.336	46.059	49.802	53.203	57.342	60.275	35
36	17.887	21.336	25.643	35.336	47.212	50.998	54.437	58.619	61.582	36
37	18.586	22.106	26.492	36.335	48.363	52.192	55.668	59.892	62.884	37
38	19.289	22.878	27.343	37.335	49.513	53.384	56.896	61.162	64.182	38
39	19.996	23.654	28.196	38.335	50.660	54.572	58.120	62.428	65.476	39
40	20.707	24.433	29.051	39.335	51.805	55.758	59.342	63.691	66.766	40
41	21.421	25.215	29.907	40.335	52.949	56.942	60.561	64.950	68.053	41
42	22.138	25.999	30.765	41.335	54.090	58.124	61.777	66.206	69.336	42
43	22.859	26.785	31.625	42.335	55.230	59.304	62.990	67.459	70.616	43
44	23.584	27.575	32.487	43.335	56.369	60.481	64.202	68.710	71.893	44
45	24.311	28.366	33.350	44.335	57.505	61.656	65.410	69.957	73.166	45
46	25.042	29.160	34.215	45.335	58.641	62.830	66.617	71.201	74.437	46
47	25.775	29.956	35.081	46.335	59.774	64.001	67.821	72.443	75.704	47
48	26.511	30.755	35.949	47.335	60.907	65.171	69.023	73.683	76.969	48
49	27.249	31.555	36.818	48.335	62.038	66.339	70.222	74.919	78.231	49
50	27.991	32.357	37.689	49.335	63.167	67.505	71.420	76.154	79.490	50



تابع الجدول (٨)
القيم الحرجة لتوزيع χ^2

ν \ α	0.995	0.975	0.9	0.5	0.1	0.05	0.025	0.01	0.005	ν
51	28.735	33.162	38.560	50.335	64.295	68.669	72.616	77.386	80.747	51
52	29.481	33.968	39.433	51.335	65.422	69.832	73.810	78.616	82.001	52
53	30.230	34.776	40.308	52.335	66.548	70.993	75.002	79.843	83.253	53
54	30.981	35.586	41.183	53.335	67.673	72.153	76.192	81.069	84.502	54
55	31.735	36.398	42.060	54.335	68.796	73.311	77.380	82.292	85.749	55
56	32.490	37.212	42.937	55.335	69.918	74.468	78.567	83.513	86.994	56
57	33.248	38.027	43.816	56.335	71.040	75.624	79.752	84.733	88.237	57
58	34.008	38.844	44.696	57.335	72.160	76.778	80.936	85.950	89.477	58
59	34.770	39.662	45.577	58.335	73.279	77.931	82.117	87.166	90.715	59
60	35.534	40.482	46.459	59.335	74.397	79.082	83.298	88.379	91.952	60
61	36.300	41.303	47.342	60.335	75.514	80.232	84.476	89.591	93.186	61
62	37.068	42.126	48.226	61.335	76.630	81.381	85.654	90.802	94.419	62
63	37.838	42.950	49.111	62.335	77.745	82.529	86.830	92.010	95.649	63
64	38.610	43.776	49.996	63.335	78.860	83.675	88.004	93.217	96.878	64
65	39.383	44.603	50.883	64.335	79.973	84.821	89.177	94.422	98.105	65
66	40.158	45.431	51.770	65.335	81.085	85.965	90.349	95.626	99.331	66
67	40.935	46.261	52.659	66.335	82.197	87.108	91.519	96.828	100.55	67
68	41.713	47.092	53.548	67.334	83.308	88.250	92.689	98.028	101.78	68
69	42.494	47.924	54.438	68.334	84.418	89.391	93.856	99.228	103.00	69
70	43.275	48.758	55.329	69.334	85.527	90.531	95.023	100.43	104.21	70
71	44.058	49.592	56.221	70.334	86.635	91.670	96.189	101.62	105.43	71
72	44.843	50.428	57.113	71.334	87.743	92.808	97.353	102.82	106.65	72
73	45.629	51.265	58.006	72.334	88.850	93.945	98.516	104.01	107.86	73
74	46.417	52.103	58.900	73.334	89.956	95.081	99.678	105.20	109.07	74
75	47.206	52.942	59.795	74.334	91.061	96.217	100.84	106.39	110.29	75
76	47.997	53.782	60.690	75.334	92.166	97.351	102.00	107.58	111.50	76
77	48.788	54.623	61.586	76.334	93.270	98.484	103.16	108.77	112.70	77
78	49.582	55.466	62.483	77.334	94.373	99.617	104.32	109.96	113.91	78
79	50.376	56.309	63.380	78.334	95.476	100.75	105.47	111.14	115.12	79
80	51.172	57.153	64.278	79.334	96.578	101.88	106.63	112.33	116.32	80
81	51.969	57.998	65.176	80.334	97.680	103.01	107.78	113.51	117.52	81
82	52.767	58.845	66.076	81.334	98.780	104.14	108.94	114.69	118.73	82
83	53.567	59.692	66.976	82.334	99.880	105.27	110.09	115.88	119.93	83
84	54.368	60.540	67.876	83.334	100.98	106.39	111.24	117.06	121.13	84
85	55.170	61.389	68.777	84.334	102.08	107.52	112.39	118.24	122.32	85
86	55.973	62.239	69.679	85.334	103.18	108.65	113.54	119.41	123.52	86
87	56.777	63.089	70.581	86.334	104.28	109.77	114.69	120.59	124.72	87
88	57.582	63.941	71.484	87.334	105.37	110.90	115.84	121.77	125.91	88
89	58.389	64.793	72.387	88.334	106.47	112.02	116.99	122.94	127.11	89
90	59.196	65.647	73.291	89.334	107.56	113.15	118.14	124.12	128.30	90
91	60.005	66.501	74.196	90.334	108.66	114.27	119.28	125.29	129.49	91
92	60.815	67.356	75.101	91.334	109.76	115.39	120.43	126.46	130.68	92
93	61.625	68.211	76.006	92.334	110.85	116.51	121.57	127.63	131.87	93
94	62.437	69.068	76.912	93.334	111.94	117.63	122.72	128.80	133.06	94
95	63.250	69.925	77.818	94.334	113.04	118.75	123.86	129.97	134.25	95
96	64.063	70.783	78.725	95.334	114.13	119.87	125.00	131.14	135.43	96
97	64.878	71.642	79.633	96.334	115.22	120.99	126.14	132.31	136.62	97
98	65.694	72.501	80.541	97.334	116.32	122.11	127.28	133.48	137.80	98
99	66.510	73.361	81.449	98.334	117.41	123.23	128.42	134.64	138.99	99
100	67.328	74.222	82.358	99.334	118.50	124.34	129.56	135.81	140.17	100

Area corresponding to percentage points

.8, .6, .4, .2, 0 — 1.0, 2.0, 3.0, 4.0 — F

	α	1	2	3	4	5	6	7	8	9	10	11	α	
1	.05	161	199	216	225	230	234	237	239	241	241	243	.05	1
	.025	648	800	864	900	922	937	948	957	963	969	973	.025	
	.01	4050	5000	5400	5620	5760	5860	5930	5980	6020	6060	6080	.01	
2	.05	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	.05	2
	.025	38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.4	.025	
	.01	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	.01	
3	.05	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	.05	3
	.025	17.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.5	14.4	14.3	.025	
	.01	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	.01	
4	.05	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93	.05	4
	.025	12.2	10.6	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.79	.025	
	.01	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	.01	
5	.05	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.71	.05	5
	.025	10.0	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.57	.025	
	.01	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.99	.01	
6	.05	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	.05	6
	.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.41	.025	
	.01	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	.01	
7	.05	5.59	4.74	4.35	4.12	3.97	3.87	3.77	3.73	3.68	3.64	3.60	.05	7
	.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.89	4.82	4.76	4.71	.025	
	.01	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	.01	
8	.05	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	.05	8
	.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.25	.025	
	.01	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.73	.01	
9	.05	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	.05	9
	.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.91	.025	
	.01	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.18	.01	
10	.05	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	.05	10
	.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.67	.025	
	.01	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	.01	

تابع الجدول (٩)

القيم الحرجة لتوزيع ف

	α	12	15	20	24	30	40	60	120	∞	α	
1	.05	244	246	248	249	250	251	252	253	254	.05	1
	.025	977	985	993	997	1000	1010	1010	1010	1020	.025	
	.01	6110	6160	6210	6230	6260	6290	6310	6340	6370	.01	
2	.05	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5	.05	2
	.025	39.4	39.4	39.4	39.5	39.5	39.5	39.5	39.5	39.5	.025	
	.01	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5	.01	
3	.05	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53	.05	3
	.025	14.3	14.3	14.2	14.1	14.1	14.0	14.0	13.9	13.9	.025	
	.01	27.1	26.9	26.7	26.6	26.5	26.4	26.3	26.2	26.1	.01	
4	.05	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63	.05	4
	.025	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26	.025	
	.01	14.4	14.2	14.0	13.9	13.8	13.7	13.7	13.6	13.5	.01	
5	.05	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36	.05	5
	.025	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02	.025	
	.01	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02	.01	
6	.05	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67	.05	6
	.025	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85	.025	
	.01	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88	.01	
7	.05	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23	.05	7
	.025	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14	.025	
	.01	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65	.01	
8	.05	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93	.05	8
	.025	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67	.025	
	.01	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86	.01	
9	.05	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71	.05	9
	.025	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33	.025	
	.01	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31	.01	
10	.05	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54	.05	10
	.025	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08	.025	
	.01	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91	.01	

تابع الجدول (٩)

القيم الحرجة لتوزيع ف

	α	1	2	3	4	5	6	7	8	9	10	11	α	
11	.05	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	.05	11
	.025	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.48	.025	
	.01	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.46	.01	
12	.05	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	.05	12
	.025	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.32	.025	
	.01	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.22	.01	
15	.05	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	.05	15
	.025	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	3.01	.025	
	.01	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	.01	
20	.05	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	.05	20
	.025	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.72	.025	
	.01	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	.01	
24	.05	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.22	.05	24
	.025	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.59	.025	
	.01	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.09	.01	
30	.05	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	.05	30
	.025	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.46	.025	
	.01	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.90	.01	
40	.05	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	.05	40
	.025	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.33	.025	
	.01	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	.01	
60	.05	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	.05	60
	.025	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.22	.025	
	.01	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	.01	
120	.05	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.87	.05	120
	.025	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.10	.025	
	.01	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.40	.01	
∞	.05	3.84	3.00	2.60	2.37	2.2[illegible]	2.10	2.01	1.94	1.88	1.83	1.79	.05	∞
	.025	5.02	3.69	3.11	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.99	.025	
	.01	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.25	.01	

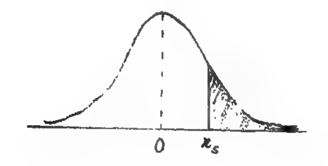
تابع الجدول (٩)

القيم الحرجة لتوزيع ف

	α	12	15	20	24	30	40	60	120	∞	α	
11	.05	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40	.05	11
	.025	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88	.025	
	.01	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60	.01	
12	.05	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30	.05	12
	.025	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72	.025	
	.01	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36	.01	
15	.05	2.48	2.40	2.33	2.39	2.25	2.20	2.16	2.11	2.07	.05	15
	.025	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40	.025	
	.01	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87	.01	
20	.05	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84	.05	20
	.025	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09	.025	
	.01	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42	.01	
24	.05	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73	.05	24
	.025	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94	.025	
	.01	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21	.01	
30	.05	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62	.05	30
	.025	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79	.025	
	.01	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01	.01	
40	.05	2.04	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51	.05	40
	.025	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64	.025	
	.01	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80	.01	
60	.05	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39	.05	60
	.025	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48	.025	
	.01	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60	.01	
120	.05	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25	.05	120
	.025	2.05	1.95	1.82	1.76	1.69	1.61	1.53	1.43	1.31	.025	
	.01	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38	.01	
∞	.05	1.75	1.67	.57	1.52	1.46	1.39	1.32	1.22	1.00	.05	∞
	.025	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00	.025	
	.01	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00	.01	

الجدول (١٠)
القيم الحرجة لمعامل ارتباط الرتب (سبيرمان)

n	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$
5	0.900	—	—	—
6	0.829	0.886	0.943	—
7	0.714	0.786	0.893	—
8	0.643	0.738	0.833	0.881
9	0.600	0.683	0.783	0.833
10	0.564	0.648	0.745	0.794
11	0.523	0.623	0.736	0.818
12	0.497	0.591	0.703	0.780
13	0.475	0.566	0.673	0.745
14	0.457	0.545	0.646	0.716
15	0.441	0.525	0.623	0.689
16	0.425	0.507	0.601	0.666
17	0.412	0.490	0.582	0.645
18	0.399	0.476	0.564	0.625
19	0.388	0.462	0.549	0.608
20	0.377	0.450	0.534	0.591
21	0.368	0.438	0.521	0.576
22	0.359	0.428	0.508	0.562
23	0.351	0.418	0.496	0.549
24	0.343	0.409	0.485	0.537
25	0.336	0.400	0.475	0.526
26	0.329	0.392	0.465	0.515
27	0.323	0.385	0.456	0.505
28	0.317	0.377	0.448	0.496
29	0.311	0.370	0.440	0.487
30	0.305	0.364	0.432	0.478



الجدول (١١)

التحويل ع = $\frac{١}{٢}$ لو $\frac{١ + ر}{١ - ر}$ لمعامل الارتباط

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
.0	0.000	0.010	0.020	0.030	0.040	0.050	0.060	0.070	0.080	0.090
.1	100	.110	.121	.131	.141	.151	.161	.172	.182	.192
.2	203	.213	.224	.234	.245	.255	.266	277	.288	299
.3	.310	.321	.332	.343	.354	.365	.377	388	400	412
.4	.424	.436	.448	.460	.472	.485	.497	510	.523	536
.5	.549	.563	.576	.590	.604	.618	.633	648	662	678
.6	.693	.709	.725	.741	.758	.775	.793	.811	829	848
.7	.867	.887	.908	.929	.950	.973	.996	1.020	1.045	1.071
.8	1.099	1.127	1.157	1.188	1.221	1.256	1.293	1.333	1.376	1.422
	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
90	1.472	1.478	1.483	1.488	1.494	1.499	1.505	1.510	1.516	1.522
.91	1.528	1.533	1.539	1.545	1.551	1.557	1.564	1.570	1.576	1.583
92	1.589	1.596	1.602	1.609	1.616	1.623	1.630	1.637	1.644	1.651
93	1.658	1.666	1.673	1.681	1.689	1.697	1.705	1.713	1.721	1.730
94	1.738	1.747	1.756	1.764	1.774	1.783	1.792	1.802	1.812	1.822
.95	1.832	1.842	1.853	1.863	1.874	1.886	1.897	1.909	1.921	1.933
.96	1.946	1.959	1.972	1.986	2.000	2.014	2.029	2.044	2.060	2.076
97	2.092	2.109	2.127	2.146	2.165	2.185	2.205	2.227	2.249	2.273
98	2.298	2.323	2.351	2.380	2.410	2.443	2.477	2.515	2.555	2.599
99	2.646	2.700	2.759	2.826	2.903	2.994	3.106	3.250	3.453	3.800

الجدول (١٢)

قيم معامل الارتباط ر بدلالة ع

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.000	0.010	0.020	0.030	0.040	0.050	0.060	0.070	0.080	0.090
.1	.100	.110	119	.129	.139	.149	.159	.168	.178	.187
.2	.197	.207	.216	.226	.236	.245	.254	.264	.273	.282
.3	.291	.300	.310	.319	.327	.336	.345	.354	.363	.371
.4	380	.389	.397	.405	.414	.422	.430	.438	.446	.454
.5	.462	470	.478	.485	.493	.500	.508	.515	.523	.530
.6	.537	.544	.551	.558	.565	.572	.578	.585	.592	.598
.7	.604	.611	.617	.623	.629	.635	.641	.647	.653	.658
.8	.664	.670	.675	.680	.686	.691	.696	.701	.706	.711
.9	.716	.721	.726	.731	.735	.740	.744	.749	.753	.757
1.0	.762	.766	.770	.774	.778	.782	.786	.790	.793	.797
1.1	.800	.804	.808	.811	.814	.818	.821	.824	.828	.831
1.2	.834	.837	.840	.843	.846	.848	.851	.854	.856	.859
1.3	.862	.864	.867	.869	.872	.874	.876	.879	881	.883
1.4	.885	.888	.890	.892	.894	.896	.898	.900	902	.903
1.5	.905	907	.909	.910	.912	.914	.915	.917	.919	.920
1.6	.922	923	.925	.926	.928	.929	.930	.932	.933	.934
1.7	.935	937	.938	.939	.940	.941	.942	.944	945	.946
1.8	.947	948	.949	.950	.951	.952	.953	954	.954	.955
1.9	.956	957	.958	.959	.960	.960	.961	.962	.963	.963
2.0	.964	965	.965	.966	.967	.967	.968	.969	.969	.970
2.1	.970	.971	.972	.972	.973	.973	.974	.974	.975	.975
2.2	.976	.976	.977	.977	.978	.978	.978	.979	.979	.980
2.3	.980	.980	.981	.981	.982	.982	.982	.983	.983	.983
2.4	.984	984	.984	.985	.985	.985	.986	986	.986	.986
2.5	.987	.987	.987	.987	.988	.988	.988	.988	.989	.989
2.6	.989	.989	.989	.990	.990	.990	.990	.990	.991	991
2.7	.991	.991	.991	.992	.992	.992	.992	.992	992	.992
2.8	.993	.993	.993	.993	.993	.993	.993	.994	.994	.994
2.9	.994	.994	.994	.994	.994	.995	.995	.995	.995	.995

الجدول (١٣)

الاحتمالات المتجمعة في اختبار التلاحقات

	a								
(n_1, n_2)	2	3	4	5	6	7	8	9	10
(2, 3)	0.200	0.500	0.900	1.000					
(2, 4)	0.133	0.400	0.800	1.000					
(2, 5)	0.095	0.333	0.714	1.000					
(2, 6)	0.071	0.286	0.643	1.000					
(2, 7)	0.056	0.250	0.583	1.000					
(2, 8)	0.044	0.222	0.533	1.000					
(2, 9)	0.036	0.200	0.491	1.000					
(2, 10)	0.030	0.182	0.455	1.000					
(3, 3)	0.100	0.300	0.700	0.900	1.000				
(3, 4)	0.057	0.200	0.543	0.800	0.971	1.000			
(3, 5)	0.036	0.143	0.429	0.714	0.929	1.000			
(3, 6)	0.024	0.107	0.345	0.643	0.881	1.000			
(3, 7)	0.017	0.083	0.283	0.583	0.833	1.000			
(3, 8)	0.012	0.067	0.236	0.533	0.788	1.000			
(3, 9)	0.009	0.055	0.200	0.491	0.745	1.000			
(3,10)	0.007	0.045	0.171	0.455	0.706	1.000			
(4, 4)	0.029	0.114	0.371	0.629	0.886	0.971	1.000		
(4, 5)	0.016	0.071	0.262	0.500	0.786	0.929	0.992	1.000	
(4, 6)	0.010	0.048	0.190	0.405	0.690	0.881	0.976	1.000	
(4, 7)	0.006	0.033	0.142	0.333	0.606	0.833	0.954	1.000	
(4, 8)	0.004	0.024	0.109	0.279	0.533	0.788	0.929	1.000	
(4, 9)	0.003	0.018	0.085	0.236	0.471	0.745	0.902	1.000	
(4, 10)	0.002	0.014	0.068	0.203	0.419	0.706	0.874	1.000	
(5, 5)	0.008	0.040	0.167	0.357	0.643	0.833	0.960	0.992	1.000
(5, 6)	0.004	0.024	0.110	0.262	0.522	0.738	0.911	0.976	0.998
(5, 7)	0.003	0.015	0.076	0.197	0.424	0.652	0.854	0.955	0.992
(5, 8)	0.002	0.010	0.054	0.152	0.347	0.576	0.793	0.929	0.984
(5, 9)	0.001	0.007	0.039	0.119	0.287	0.510	0.734	0.902	0.972
(5, 10)	0.001	0.005	0.029	0.095	0.239	0.455	0.678	0.874	0.958
(6, 6)	0.002	0.013	0.067	0.175	0.392	0.608	0.825	0.933	0.987
(6, 7)	0.001	0.008	0.043	0.121	0.296	0.500	0.733	0.879	0.966
(6, 8)	0.001	0.005	0.028	0.086	0.226	0.413	0.646	0.821	0.937
(6, 9)	0.000	0.003	0.019	0.063	0.175	0.343	0.566	0.762	0.902
(6, 10)	0.000	0.002	0.013	0.047	0.137	0.288	0.497	0.706	0.864
(7, 7)	0.001	0.004	0.025	0.078	0.209	0.383	0.617	0.791	0.922
(7, 8)	0.000	0.002	0.015	0.051	0.149	0.296	0.514	0.704	0.867
(7, 9)	0.000	0.001	0.010	0.035	0.108	0.231	0.427	0.622	0.806
(7, 10)	0.000	0.001	0.006	0.024	0.080	0.182	0.355	0.549	0.743
(8, 8)	0.000	0.001	0.009	0.032	0.100	0.214	0.405	0.595	0.786
(8, 9)	0.000	0.001	0.005	0.020	0.069	0.157	0.319	0.500	0.702
(8, 10)	0.000	0.000	0.003	0.013	0.048	0.117	0.251	0.419	0.621
(9, 9)	0.000	0.000	0.003	0.012	0.044	0.109	0.238	0.399	0.601
(9, 10)	0.000	0.000	0.002	0.008	0.029	0.077	0.179	0.319	0.510
(10, 10)	0.000	0.000	0.001	0.004	0.019	0.051	0.128	0.242	0.414

تابع الجدول (١٣)
الاحتمالات المتجمعة في اختبار التلاحقات

(n_1, n_2)	a									
	11	12	13	14	15	16	17	18	19	20
(2, 3)										
(2, 4)										
(2, 5)										
(2, 6)										
(2, 7)										
(2, 8)										
(2, 9)										
(2, 10)										
(3, 3)										
(3, 4)										
(3, 5)										
(3, 6)										
(3, 7)										
(3, 8)										
(3, 9)										
(3, 10)										
(4, 4)										
(4, 5)										
(4, 6)										
(4, 7)										
(4, 8)										
(4, 9)										
(4, 10)										
(5, 5)										
(5, 6)	1.000									
(5, 7)	1.000									
(5, 8)	1.000									
(5, 9)	1.000									
(5, 10)	1.000									
(6, 6)	0.998	1.000								
(6, 7)	0.992	0.999	1.000							
(6, 8)	0.984	0.998	1.000							
(6, 9)	0.972	0.994	1.000							
(6, 10)	0.958	0.990	1.000							
(7, 7)	0.975	0.996	0.999	1.000						
(7, 8)	0.949	0.988	0.998	1.000	1.000					
(7, 9)	0.916	0.975	0.994	0.999	1.000					
(7, 10)	0.879	0.957	0.990	0.998	1.000					
(8, 8)	0.900	0.968	0.991	0.999	1.000	1.000				
(8, 9)	0.843	0.939	0.980	0.996	0.999	1.000	1.000			
(8, 10)	0.782	0.903	0.964	0.990	0.998	1.000	1.000			
(9, 9)	0.762	0.891	0.956	0.988	0.997	1.000	1.000	1.000		
(9, 10)	0.681	0.834	0.923	0.974	0.992	0.999	1.000	1.000	1.000	
(10, 10)	0.586	0.758	0.872	0.949	0.981	0.996	0.999	1.000	1.000	1.000

الجدول (١٤)

القيم الحرجة لاختبار ويلكوكسن للمقارنات التزاوجية

n	One-sided $\alpha = 0.01$ Two-sided $\alpha = 0.02$	One-sided $\alpha = 0.025$ Two-sided $\alpha = 0.05$	One-sided $\alpha = 0.05$ Two-sided $\alpha = 0.10$
5			1
6		1	2
7	0	2	4
8	2	4	6
9	3	6	8
10	5	8	11
11	7	11	14
12	10	14	17
13	13	17	21
14	16	21	26
15	20	25	30
16	24	30	36
17	28	35	41
18	33	40	47
19	38	46	54
20	43	52	60
21	49	59	68
22	56	66	75
23	62	73	83
24	69	81	92
25	77	90	101
26	85	98	110
27	93	107	120
28	102	117	130
29	111	127	141
30	120	137	152

† Reproduced from F. Wilcoxon and R. A. Wilcox *Some Rapid Approximate Statistical Procedures*, American Cyanamid Company, Pearl River, N.Y., 1964 by permission of the American Cyanamid Company

الجدول (١٥)

الاحتمالات ل (س ⩽ س٠) فى اختبار الاتجاه

For $n = 3$, $F(2) = 1 - 0.167 = 0.833$.
For $n = 4$, $F(3) = 1 - 0.375 = 0.625$, $F(4) = 1 - 0.167 = 0.833$, etc.

x	n=3
	0.
0	167
1	500

x	n=4
	0.
0	042
1	167
2	375

x	n=5
	0.
0	008
1	042
2	117
3	242
4	408

x	n=6
	0.
0	001
1	008
2	028
3	068
4	136
5	235
6	360
7	500

x	n=7
	0.
1	001
2	005
3	015
4	035
5	068
6	119
7	191
8	281
9	386
10	500

x	n=8
	0.
2	001
3	003
4	007
5	016
6	031
7	054
8	089
9	138
10	199
11	274
12	360
13	452

x	n=9
	0.
4	001
5	003
6	006
7	012
[illegible]	022
9	038
10	060
11	090
12	130
13	179
14	238
15	306
16	381
17	460

x	n=10
	0.
6	001
7	002
8	005
9	008
10	014
11	023
12	036
13	054
14	078
15	108
16	146
17	190
18	242
19	300
20	364
21	431
22	500

x	n=11
	0.
8	001
9	002
10	003
11	005
12	008
13	013
14	020
15	030
16	043
17	060
18	082
19	109
20	141
21	179
22	223
23	271
24	324
25	381
26	440
27	500

x	n=20
	0.
50	001
51	002
52	002
53	003
54	004
55	005
56	006
57	007
58	008
59	010
60	012
61	014
62	017
63	020
64	023
65	027
66	032
67	037
68	043
69	049
70	056
71	064
72	073
73	082
74	093
75	104
76	117
77	130
78	144
79	159
80	176
81	193
82	211
83	230
84	250
85	271
86	293
87	315
88	339
89	362
90	387
91	411
92	436
93	462
94	487

x	n=19
	0.
43	001
44	002
45	002
46	003
47	003
48	004
49	005
50	006
51	008
52	010
53	012
54	014
55	017
56	021
57	025
58	029
59	034
60	040
61	047
62	054
63	062
64	072
65	082
66	093
67	105
68	119
69	133
70	149
71	166
72	184
73	203
74	223
75	245
76	267
77	290
78	314
79	339
80	365
81	391
82	418
83	445
84	473
85	500

x	n=18
	0.
38	001
39	002
40	003
41	003
42	004
43	005
44	007
45	009
46	011
47	013
48	016
49	020
50	024
51	029
52	034
53	041
54	048
55	056
56	066
57	076
58	088
59	100
60	115
61	130
62	147
63	165
64	184
65	205
66	227
67	250
68	275
69	300
70	327
71	354
72	383
73	411
74	441
75	470
76	500

x	n=17
	0.
32	001
33	002
34	002
35	003
36	004
37	005
38	007
39	009
40	011
41	014
42	017
43	021
44	026
45	032
46	038
47	046
48	054
49	064
50	076
51	088
52	102
53	118
54	135
55	154
56	174
57	196
58	220
59	245
60	271
61	299
62	328
63	358
64	388
65	420
66	452
67	484

x	n=16
	0.
2.	001
28	002
29	002
30	003
31	004
32	006
33	008
34	010
35	013
36	016
37	021
38	026
39	032
40	039
41	048
42	058
43	070
44	083
45	097
46	114
47	133
48	153
49	175
50	199
51	225
52	253
53	282
54	313
55	345
56	378
57	412
58	447
59	482

x	n=15
	0.
23	001
24	002
25	003
26	004
27	006
28	008
29	010
30	014
31	018
32	023
33	029
34	037
35	046
36	057
37	070
38	084
39	101
40	120
41	141
42	164
43	190
44	218
45	248
46	279
47	313
48	349
49	385
50	423
51	461
52	500

x	n=14
	0.
18	001
19	002
20	002
21	003
22	005
23	007
24	010
25	013
26	018
27	024
28	031
29	040
30	051
31	063
32	079
33	096
34	117
35	140
36	165
37	194
38	225
39	259
40	295
41	334
42	374
43	415
44	457
45	500

x	n=13
	0.
14	001
15	001
16	002
17	003
18	005
19	007
20	011
21	015
22	021
23	029
24	038
25	050
26	064
27	082
28	102
29	126
30	153
31	184
32	218
33	255
34	295
35	338
36	383
37	429
38	476

x	n=12
	0.
11	001
12	002
13	003
14	004
15	007
16	010
17	016
18	022
19	031
20	043
21	058
22	076
23	098
24	125
25	155
26	190
27	230
28	273
29	319
30	369
31	420
32	473

References

Bhattacharyya, G. and Johnson R.: Statistical Concepts and Methods, Wiley, 1977.

Bishop, O. N.: Statistics for Biology, Longman, 1971.

Cochran, W. G.: Sampling Techniques, Wiley, 1977.

Cochran, W. G. and Cox, G. M.: Experimental Designs, Wiley, 1957.

Colquohoun, D.: Lectures on Biostatistics, Clarendon Press, Oxford, 1971.

Freund, J. E.: Modern Elementary Statistics, Pentice - Hall, 1987.

Gunst, R. E. and Mason, R.L.: Regression Analysis and Application.

Hays, W. L.: Statistics for Social Sciences, Holt, Rinehart and Winston, 1973.

Hill, A. B.: A Short Textbook of Medical Statistics, Lancet, 1984.

Kreyszig, E.: Introductory Mathematical Statistics, Wiley, 1970.

Krumben, W. C. and Graybill, F. A.: Statistical Models in Geology, McGraw-Hill, 1965.

Milton, J. S. and Others: Introduction to Statistics, Heath and Company, 1986.

Nalimov, V. V.: The Application of Mathematical Statistics to Chemical Analysis.

Skane, D. H.: Elementary Statistics, Addison-Wesley, 1985.

Sokal, R. R. and Rohlf, F. J.: Introduction to Biostatistics, Freeman, 1973.

Snedecor, G. W. and Coehran, W. G.: Statistical Methods, Iowa State Univ. Press, 1980.

Sykes, M. N. and Vickers, M. D.: Principles of Clinical Measurement, McGraw-Hill, 1982.

Walpole, R. and Myers, R.: Probability and Statistics for Engineers and Scientists, Macmillan, 1978.

Wardlaw, A.C: Practical Statistics for Experimental Biologists.

Winer, B. J.: Statistical Principles in Experimental Designs, McGraw-Hill, 1971.

Zuwaylif, F. H.: Applied Business Statistics, Addison-Wesley, 1984.

يعرض هذا الكتاب شيئا مما يسمى بالإحصاء التطبيقي . وهو يتناول جل المفاهيم والطرق ، ونماذج الإحصائية التي يحتاج إليها الطلاب والباحثون التطبيقيون في مختلف ميادين البحث العلمي . وتستند معالجة الموضوعات المقدمة على أمثلة توضيحية تيسر للقارئ تقبل ما يعرض من مادة وتضيء له طريق استخدامها في التطبيقات العملية . وتتدعم هذه المعالجة بتمارين صممت لتستثير فكر القارئ وتعاونه على ربط النقاط الأساسية فيما يقرأ ، كما تمنحه الفرصة لتقويم ما استوعبه من المادة وتحسين هذا الاستيعاب عن طريق التغذية المرتجعة التي توفرها الأجوبة الشاملة المعطاه بالملحق (١) لهذا الكتاب .

والناشر إذ يفخر بتقديم هذا المرجع القيم يرجو أن يكون فيه إضافة جوهرية للمكتبة العربية .

